DOCUMENT RESUME

ED 220 539 TM 820 570

AUTHOR Burry, James; And Others
TITLE Testing in the Nation's Schools and Districts: How Much? What Kinds? To What Ends? At What Costs?
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO CSE-R-194
PUB DATE Mar 82
NOTE 142p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Cost Effectiveness; *Educational Assessment; Educational Economics; Educational Policy; Minimum Competency Testing; National Surveys; Program Costs; *School Districts; Testing; *Testing Programs; Test Results; *Test Use

ABSTRACT

The Center for the Study of Evaluation's (CSE) Test Use Project (1979) has gathered information that is nationally representative and illustrative of the entire range of tests being administered. The primary intention of this phase of the ongoing study is to identify the direct and indirect costs of testing. The four papers included here offer school districts a fresh vantage point from which to consider how their assessment programs can be improved to meet a variety of decision audiences. Bruce Choppin discusses the survey's sampling procedures and offers an overview of the main findings, concluding with ideas to reduce the amount of testing time while maintaining its relevance for various audiences. Donald Dorr-Bremme amplifies the initial findings in a teacher-as-practical-decision-maker context, with implications for the design and implementation of future assessment programs. James Burry discusses CSE's test use findings indicating teachers' stated uses of assessment information for classroom decisions and recommending methodological, technical, and organizational considerations to be addressed to produce more efficient assessment programs. James Catterall discusses cost-accounting, cost-effectiveness, and cost-benefit paradigms, and offers a theoretical model for thinking about costs and testing. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220539

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TESTING IN THE NATION'S SCHOOLS AND DISTRICTS: HOW MUCH? WHAT KINDS? TO WHAT ENDS? AT WHAT COSTS?

James Burry
James Catterall
Bruce Choppin
Donald Dorr-Bremme

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.S. Cizy
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CSE Report No. 194
1982

The papers in this volume were originially presented in a symposium at the Annual Meeting of the American Educational Research Association, New York, 1982.

TM 820 570

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

TABLE OF CONTENTS

	Page
Introduction	i
How Schools Make Use of Test Results Bruce Choppin	1
Assessing Students: Teachers' Routine Practices and Reasoning Donald Dorr-Bremme	18
Using Tests: Who Do We Believe and What Does it Mean? James Burry	52
The Cost of School Testing Programs James Catterall	96

INTRODUCTION

CSE's Test Use Project has been gathering information bearing on a range of testing issues for students, teachers, administrators, researchers, and policy makers. It is clear that our schools do a great deal of student achievement testing, and some limited information has already been collected on certain practices affecting our students in some areas of the country. Until the CSE study, however, we have lacked information that is nationally representative and illustrative of the entire range of tests being administered, and yet which is sufficiently focused to be of use in test-based policy matters.

CSE has been concerned, first, that there is a lack of descriptive data reflecting the entire testing picture--the range of tests being administered, their associated users and consumers, and the range of students affected by particular kinds of tests. Second, there is also a lack of the more inferential utilization data--the primary and secondary users of test information, the intended and actual uses of test information, variations in use across users and organizational settings, the kinds of decisions made on the basis of test information, the kinds of students thereby affected, and the attendant costs of the testing enterprise.

Since the inception of the Test Use Project in December 1979, we have been examining these kinds of issues in a broad framework which defines testing to include formal tests, both norm- and criterion-referenced; curriculum-embedded measures; district-, school; and teacher-developed tests; as well as the more informal measures such as teacher quizzes,

observations, and other interactions with students. In short, our study has not aimed at any single kind of test, user, or student. But the study is also sharply focused in this broad framework, and examines some of the more troublesome aspects of testing: student achievement testing in language arts and mathematics; at selected grade levels where testing may critically affect large numbers of students and their teachers--fourth and sixth grades in elementary schools and tenth grade in high schools. Finally, information on these matters has been primarily reported to us by teachers and principals--those who are closely involved in the use of tests.

The Test Use Project has been proceeding in two overlapping phases. Phase I, taking place between December 1979 and November 1981, led to the collection and analyses of survey data from a national sample of teachers and principals representing the targeted grades/schools. During Phase II of the study, which began in February 1981 and will conclude in November 1982, the project is conducting on-site studies in a small number of schools. The primary intention of this phase of the study is to identify the direct and indirect costs of testing.

The four papers in this report were first presented in an AERA symposium on test use in New York, 1982. Each of the papers derives from CSE fieldwork conducted to inform the national survey design and from data collected in that survey and in current examination of the costs of testing.

Beginning the report, Choppin discusses the survey's sampling procedures and offers an overview of some of the main findings: how much testing is taking place, with what kinds of tests, how they are used, and

their role in teachers' decision making. He concludes with ideas about how to reduce the amount of testing time while maintaining its relevance for various audiences.

Dorr-Bremme amplifies some of the initial findings and presents them in a context which views the teacher as practical decision maker. This view of the teacher has implications for the design and implementation of assessment programs in the future.

Burry places CSE's test use findings in the context of previous studies of the phenomenon and relates them to other relevant literature. He draws implications and recommendations reflecting methodological, technical, and organizational considerations to be addressed before more efficient assessment programs are considered.

Finally, Catterall's paper provides an inquiry into the costs of testing by discussing cost-accounting, cost-effectiveness, and cost-benefit paradigms, and offers an economy of information perspective as a theoretical model for thinking about costs and testing.

Taken together, the four papers in the report offer schools and districts a fresh vantage point from which to consider how their assessment programs can be improved to meet a variety of decision audiences.

HOW SCHOOLS MAKE USE OF TEST RESULTS

Bruce Choppin

INTRODUCTION

Although the literature contains much information on teachers' attitudes to tests and testing, and on the use of specific tests, there is very little published regarding the scale of the total testing enterprise. It is generally recognized that testing plays an important role in schooling within the United States--and the impression of educationists in other countries is that more testing is conducted here than anywhere else--but finding evidence about precisely how much testing is done, what sort of testing, and what use is made of the results has been difficult. Hence CSE's decision to conduct this national survey.

It was clearly not practical to try to include all grade levels and all subject areas within a study such as this, so we decided to concentrate on the basic skills areas, reading and mathematics, in the upper elementary grades, and on language arts and mathematics at the 10th grade.

SAMPLES

The sampling procedures employed were complex. We needed to obtain a nationally representative picture of the uses of testing and had only limited resources to accomplish this. Teachers were the primary target of

the survey because they conduct most of the achievement testing and are, therefore, in the best strategic position from which to judge the relevance of testing programs to their own needs. In addition, and in order to collect information on relevant contextual variables, the principals of the selected schools and district testing officers were also included in the study.

We drew a probability sample of 114 school districts from the 13,815 listed on a commercial data base using five stratifying variables: geographical region, locale, socioeconomic status of the area, the size of the school district, and policy with regard to minimum competency testing. Details are to be found in Table 1.

These five stratifying variables jointly define a 900 cell matrix, but when the population of school districts is distributed among them, 544 of the cells are found to be empty. Thus, the sampling strategy required the choosing of 114 school districts from among the remaining 356 cells. We employed a lattice sampling technique to select cells from the matrix, and then simple random sampling to select districts within cell.

Extensive telephone interviews were conducted with the officials responsible for testing and assessment within each selected school district in order to establish what the local policies in these areas were. Information was also collected which permitted us to sample two high schools and two elementary schools in each district.

The principals of the selected schools were contacted, and were sent a questionnaire to complete themselves and questionnaires for four of their teachers. In the case of elementary schools, principals were given

Table 1

Stratification Employed to Select Sample of School Districts

Stratification Variable	Categories	No. of Districts in Total Population	% of Total Enrollment in Category	No. of Responding Districts in Sample
Status on Minimum Competency Testing	MCT not required for graduation or promotion (no local option)	2703	19	22
	MCT not required but there are local options	2065	13	17
	MCT required for graduation and/or promotion (no local options)	980	18	21
	MCT required for graduation and/or promotion with local options	1778	16	16
	No MCT program mandated in 1981 at the state level	6289	34	15
Size of School District	Enrollment less than 5000	12061	37	19
	Enrollment 5000 - 9999	1059	18	22
	Enrollment 10,000 - 24,999	514	18	22
	Enrollment 25,000 - 44,999	105	8	9
	Enrollment greater than 45,000	76	19	19
SES of Area (Orshansky Index)	Wealthiest	1907	16	15
	Middle group	9051	69	61
	Poorest	2857	15	15
Geographic Region	North East	2718	25	22
	South East	1736	24	28
	Middle	5279	27	22
	West	4092	25	19
Locale	Central City	915	31	33
	Urban Fringe	3354	32	27
	Non-metropolitan	9546	37	31

instructions for sampling two 4th grade and two 6th grade teachers, but were told how to substitute 5th grade teachers if, for some reason, the quota for 4th or 6th grade teachers could not be met. At the high school level, principals were told how to draw samples of two 10th grade English teachers and two 10th grade mathematics teachers. The sampled teachers were requested to complete a detailed questionnaire about their use of tests with the chosen class.

We deliberately undersampled two large strata: those districts with enrollments less than 5,000, and those with no MCT program. This increased the possibilities for analysis within the other levels, while differential weighting would still allow the calculation of unbiased estimates of the national characteristics. In the event, it turned out that the rate of return from the largest enrollment category was lower than that from the others, so that the weighting was adjusted to correct for this. Rates of return from the four regions were also not uniform, with the southeast states having the highest rate. Again weighting factors solved the problem.

Although we obtained data from 91 of the selected school districts (rather more than 80 percent of the target figure) the rate of return from the principals and teachers was only about 60 percent. We are, therefore, less confident about generalizing to the national population than we would like to be. It also became clear that a substantial number of 5th grade teachers had been included in the elementary school sample and, since a preliminary analysis revealed no significant differences between the patterns of response between 4th, 5th, and 6th grade teachers, it was

decided to pool these. As a consequence of this, we report results only for "elementary teachers" rather than for each grade separately.

The rest of this paper is devoted to a brief overview of some of the main findings to emerge from the survey; the later papers explore selected areas in more detail. It should perhaps be pointed out that despite the modest size of our sample, the complexity of the data collected is such that we do not expect to exhaust the possibilities for useful analysis for a considerable time to come.

HOW MUCH TESTING IS TAKING PLACE?

Tables 2 and 3 summarize the results of the survey as far as the total sample is concerned. Note that at the elementary grades each class experiences about 10 hours of reading tests and about 12 hours of mathematics tests during the course of the year. This amounts amounts to about 5 percent of the total instructional time in those subjects.

In high schools we find a different picture. Tenth grade classes spend about twice as much time taking tests in these basic skill areas. they occur more frequently--rather more than once each week. The overall impact is thus rather more than 10 percent of the total available instructional time for the class.

We asked the teachers to distinguish between: (a) testing they were mandated to carry out to fulfill state requirements; (b) tests that were required by district policy; and (c) other tests given at the teachers' initiative or as part of the school assessment policy.

Table 2

Time Devoted to Testing in Typical Classes

	Total Amount of Class Time Spent on Testing per Annum	No. of Test Sessions for Typical Student	Average Length of Session
Elementary School (Grades 4-6)			
--Reading Tests	9 hrs. 56 min.	22	27 min.
--Mathematics Tests	12 hrs. 28 min.	23	32 min.
10th Grade English Class	26 hrs. 34 min.	49	32 min.
10th Grade Mathematics Class	24 hrs. 18 min.	45	33 min.

Table 3

Time Devoted to Required Testing,
As a Percentage of Total Testing Time
For Typical Classes

	Percentage Time on Testing Required by State	Percentage Time on Testing Required by Local School District	Percentage Testing Time Devoted to Non-Required Tests
Elementary School (Grades 4-6)			
--Reading	30	29	41
--Mathematics	21	25	54
10th Grade English Class	12	13	74
10th Grade Mathematics Class	9	14	77

As was to be expected, most time was spent on tests which fell in the third category, but note in Table 3 the differences between elementary and high school patterns. State requirements play a significantly larger role in the testing of reading in the elementary grades.

WHAT TESTS ARE USED?

Our initial attempts to catalogue the full range of tests being used by the teachers who fell in our sample was abandoned because of the immense size of the task. Many teachers listed as many as ten different tests or series of tests that they used with a single class and there appeared to be no individual test that was used in a majority of the schools that formed our sample. Instead, we have settled for a simple categorization which is laid out in Table 4, and which first shows minimum competency tests administered as a part of state education policy and designed either locally or at a state level. Tests which are included with curriculum materials (for instance, unit/chapter, end-of-book, or diagnostic tests), appear next, followed by commercially published tests, particularly standardized tests. The last two categories are for locally developed tests adopted at the district level and for the teachers' own tests or other tests developed within the school.

It is this last category of test, the one developed within the school itself, and usually by the teacher concerned, that takes the greatest proportion of the total time devoted to testing. This is especially true

Table 4

Types of Test Used,
As a Percentage of the Total Time
Devoted to Testing

TYPE OF TEST	Elementary Teachers		10th Grade English Teachers	10th Grade Mathematics Teachers
	Reading	Math		
Tests which form part of a statewide assessment program	3	3	5	1
Required Minimum Competency Tests	1	2	1	1
Tests included with curriculum materials	28	35	8	17
Other commercially published tests	17	18	6	3
Locally developed and district adopted tests	13	8	5	2
School or teacher developed tests	37	35	74	76

at the high school level, where three-quarters of all the testing appears to be of this type. Apart from this, it is notable that the tests included with curriculum materials appear to play a prominent role in mathematics classes.

The total amount of time devoted to statewide assessment programs and required minimum competency tests appears small. The figures presented in this table are averaged across all the teachers in our survey including those in states without any MCT program, but even if the analysis is restricted to those states where minimum competency tests are used, the proportion of time spent on them is still small.

Where minimum competency tests are required, less than 3 percent of the testing time in the elementary schools and 2 percent of the testing time in secondary schools is taken up with these tests. Where MCTs are available,but not required, they absorb less than 1 percent of the total testing time.

The picture with regard to statewide assessment programs is similar. For example, they absorb no more than about 3 percent of the total testing time at the elementary level (or about 45 minutes on average per year for reading and mathematics combined). At the high school level, 10th grade English assessment programs absorb an average of 75 minutes and mathematics programs, on average, 30 minutes. It is clear that the impact of these programs on school instruction cannot be fairly judged in terms of the additional testing burden they impose which competes for regular class time with instruction itself. Rather, as we shall see, the impact is to be

measured by the pressures that teachers report concerned with the need to preparestudents for these tests.

HOW ARE TESTS USED

All schools use tests to a greater or lesser extent. Teachers in the United States use routine testing for three main purposes: to motivate students to study harder; to provide themselves and the students with feedback about the success or failure of recent learning; and to provide some quantitative data-base for generating grades. Of course the second and third of these activities fuel the first. It is the explicit link between the testing and the subsequent feedback and grades that motivates the students to study harder. Teachers all around the world use tests for these same purposes, although the balance between the different types of feedback offered, and the importance attached to grades, varies from culture to culture. American teachers, in contrast to those elsewhere, tend to emphasize the importance of grades.

For those tests which teachers said they were required to give, either by their school district or state policy (and for brevity, I shall refer to these as mandated tests from now on), the test scripts themselves are typically sent on to the school district or state authority as appropriate. Remember that these tests absorb about one-half of the total testing time in the elementary grades and one-quarter of the total by grade 10. Of course the teacher may make some some direct use of these results before they are turned in, but an important question for us was whether or

not the teachers believed that the results were used higher up the administrative pyramid.

We asked the teachers a number of questions about the use of test data by their school authorities and the results are summarized in Table 5.

At the elementary level it seems that most principals do use test scores to identify topics that need extra emphasis, and that they follow this up with some sort of check on the teachers' response (by observing classes, by reviewing the teachers' plans, or by having the teacher write specific reports). At the secondary level, this is less frequent, but is something that the majority of the teachers say happens at least sometimes.

Almost 90 percent of the elementary teachers and about two-thirds of the secondary teachers reported that some test scripts were turned over directly to the district. However, there is a considerable difference between the reported experience of elementary and 10th grade teachers in respect of these tests. More than half the elementary teachers agreed that the results of these tests were returned to them soon enough so that they could use them to modify instruction for some or all of the students in the class, and four-fifths of these teachers said that the format in which the test results were returned was useful. By contrast, only a third of the secondary teachers reported that the test results came back soon enough to be useful and 45 percent of them stated that the result format used gave them little useful information. Seventeen percent of the secondary teachers who sent test scripts to their school district claimed that the district did not return the results at all.

Table 5

Teachers' Reports on the Extent to Which the School Makes Use of Test Results

USE OF RESULTS	Percentage of teachers reporting that the activity					
	happens routinely		occurs sometimes but not often		does not happen at all	
My principal (or the school administration) ...	Elementary Teacher	10th Grade Teacher	Elementary Teacher	10th Grade Teacher	Elementary Teacher	10th Grade Teacher
... reviews test scores to identify skill or content areas that need extra emphasis.	38	13	49	51	13	36
... checks that I am emphasizing the areas identified by test scores needing it.	32	22	48	50	20	28
... requires me to turn in the scores or grades on the tests that I routinely give my classroom.	18	11	18	18	64	71
... evaluates my teaching on the basis of test scores and/or establishes specific test-score goals for my students and me to meet.	5	2	23	14	72	84

Finally, note that according to Table 5, more than a quarter of the elementary teachers feel that they are evaluated in terms of the test scores of their students. This is almost certainly an inappropriate use of test scores and a poor way to approach such evaluation. However, the teachers may be unduly sensitive in this area. Our survey suggests that elementary school principals in general do not regard test scores as playing any significant role in teacher evaluation.

DO TESTS HELP THE TEACHER MAKE DECISIONS?

What of the decisions that teachers themselves need to make during the course of a school year? We asked the teachers to rate the importance of different sources of information, such as: scores on various types of tests; their own direct observations of students; their previous experience of teaching; and comments, reports, and grades received from previous teachers. We asked teachers (a) about decisions they made in planning their courses at the beginning of the year, (b) about the initial grouping of students,(c) about moving students from one group to another during the course of a year, or providing remedial or accelerated work, and (d) about decisions concerning the students' report card grades.

Burry's paper later in this report will explore these results in much more depth, but two general findings emerge.

The first is that for both elementary and secondary teachers, the teachers of reading and those of math, and for all four types of decision, there is a common and consistent pattern. The teachers give most weight to

their own observations and to the students' class work. Next in importance come the tests that the teachers themselves have composed. Third come tests provided with the curriculum materials. These consistently come out ahead of scores on standardized tests, district continuum or minimum competency tests, statewide assessment tests, etc.

The second finding is that while this pattern is consistent, the differences in the weights accorded to the different forms of evidence are comparatively small for decisions concerning initial planning, placement, and grouping of students. For these decisions all sources of information listed were rated as at least fairly important. However, for students' final grades the determining factors were clearly the teachers' own observations, student classwork, and the results of the teachers' own tests. The other types of information were far less important. This would seem to suggest that despite the teachers' expressed belief in, and respect for, the high quality of commercially published tests and tests originating at the district level or above, they also have a high regard for their own competence as testers. It is also reassuring that they put more faith in their own observations than in any particular test score.

REDUCING THE TIME SPENT ON TESTING

While the primary purpose of this paper has been to provide an overview of the survey results, I will conclude it with some general remarks.

The substantial amount of testing that goes on in our schools can be divided into two main categories. The first comprises the testing that is organized and executed by the individual teacher with the primary purposes of motivating the students and generating grades for them. The second is the category of mandated testing which covers all those activities required by school district or state policy which are aimed at evaluating the effectiveness of the core educational system. The quantitative data on performance developed from these tests has potential for decision making at levels above the individual classroom.

For the most part, the non-mandated testing that teachers organize and run by themselves appears to be working well. Teachers clearly put considerable trust in the results of their own tests, and make extensive use of them.

The functioning of mandated testing appears in general less satisfactory. There is room for discussion about the extent to which this effectively serves current policy requirements, and in places there is room for doubt that the scores from such testing are used intelligently (or even used at all).

One way of increasing the overall efficiency of the schools might be to reduce the total time devoted to tests thereby releasing some additional time for regular instruction. In our data there is no evidence that teachers would wish, in general, to reduce the time they spend giving their own tests, but at the moment these tests serve the teacher's own needs, but not those of policymakers at the district and state level. If there is

to be progress, perhaps it lies in the direction of the making the teachers' own tests more useful to the policymakers so that separate programs of mandated testing could be reduced or abolished. One approach to this might be to give teachers access to calibrated item banks, especially if this were combined with schoolwide or districtwide record keeping systems that kept track of all student test data. The information necessary for school, district, or state reporting could then be extracted from existing records without the need for additional testing sessions. If this information was to be credible, then teachers would need to be convinced that test scores were not being used to evaluate their own performance (a step that I would advocate in any event).

Item banks of the scope needed to make this type of scheme function are being developed. In a few districts (Portland, Oregon and Los Angeles County come to mind), they are already operational. A more urgent priority now is the development of effective data banking systems within schools that would facilitate the aggregation and interpretation of test data for the purposes suggested above. The current invasion of our schools by micro- and minicomputers suggests that solutions to the technical aspects of this problem are now available, but the design of an effective "comprehensive information center" for schools will be no easy task.

ASSESSING STUDENTS: TEACHERS' ROUTINE PRACTICES AND REASONING

Donald W. Dorr-Bremme

INTRODUCTION

American educational organizations (schools, school districts, etc.) have been called "loosely-coupled systems" (c.f. Deal, 1979, Meyer & Rowan, 1978, Montjoy & O'Toole, 1979). Schooling in the United States has been described as "pre-industrial--a cottage industry" (Dawson, 1977). And teachers in classrooms have been likened to "street-level bureaucrats" (e.g., Weatherly & Lipsky, 1977). These metaphors call attention to the relative autonomy of the classroom teacher in a multi-leveled, decision-making hierarchy, a hierarchy in which participants at each level have interests and concerns that only partially overlap, only sometimes coincide. In such a system, innovation tends to be more enduring not when it is imposed from the top down, not when it is generated from the bottom up, but when it is planned and implemented conjointly by participants at all levels (Berman & McLaughlin, 1978).

All this bears on the development and implementation of testing programs. It suggests that if those who choose testing programs and/or develop tests want those programs and tests to be useful for teachers and used in classrooms, they must (at the very least) take into account teachers' perspectives on the assessment of student achievement.

But what are teachers' perspectives on the assessment of student achievement? How do teachers think and reason about evaluating students'

performance and progress? What methods, what processes and tools, do they routinely employ in making sense of how students are doing academically? Up until now, there has been little systematically gathered information to answer such questions and the few studies that have asked them have focused on teachers' attitudes and practices with regard to standardized tests (e.g., Airasian, 1979l Airasian, Kelleghan, Madaus & Pedulla, 1977; Goslin, 1967; Resnick, 1981; Stetz & Beck, 1979. Also refer to Burry elsewhere in this report). Through the last two years, however, CSE has gathered and analyzed data on teachers' attitudes toward and uses of a broad range of types of tests and other assessment techniques. This paper reports some of those findings. More specifically, it (1) presents an analysis of teachers' routine thinking and practices in assessing students, then (2) outlines some implications of that analysis for the development of testing policy and programs, especially at the local level, i.e., in schools and school districts.

THE DATA BASE

The findings discussd here are based on data gathered in two ways.

- During the CSE test use project's first year, comprehensive semi-structured interviews were conducted with 80 educators in nine schools, three each in three school districts located in different states and geographic regions of the country. The districts and the elementary and secondary schools visited varied in size and demographic setting. Each of the interviews lasted between a half-hour and an hour and focused on assessment in the basic skills areas, reading/English, language arts and mathematics. Included among the interview respondents were 44 classroom teachers (22 elementary, 22 high school) as well as elementary school instructional specialists, high school math and department chairpersons, counselors, principals, and other school administrators. Their remarks were tape recorded, transcribed, and coded using inductively developed categories.

- During the project's second year, quesionnaires were mailed to teachers and principals in a nationally representative sample of school districts and schools. Some 486 upper elementary grade teachers and 365 high school English and meth teachers responded to this survey. (See Choppin elsewhere in this report for fuller details on the survey methods.)

I also refer in passing to data collected in an earlier CSE study of testing and test use (Yeh, 1978) conducted via self-administered questionnaires in 19 schools in five California school districts. Some 256 questionnaires were returned by teachers in grades K-6 in this study and the data they produced were reanalyzed in the process of planning for the national survey.

The findings from the national survey and from the on-site interviews are completely consonant, even though they derive from data that were gathered using entirely different elicitation frameworks. In the following discussion, I interweave the survey and interview findings, drawing upon their mutually complementary strengths.

THE FINDINGS: HOW TEACHERS ROUTINELY THINK AND ACT IN ASSESSING STUDENT ACHIEVEMENT

I turn now to the question, how do teachers routinely think and act in assessing student achievement? In answer to that question, the findings of the CSE test use project suggest that teachers think and act as practical reasoners and decision makers. That is, as they go about the business of determining how the students in their class(es) are doing:

- They orient their activities to the practical tasks they have to accomplish in their everyday routines and do so in light of the practical contingencies and exigencies that they face.

> ° And, as they do, they make sense of student's academic performances clinically. They take into account all the "data" at hand "in this particular situation." Then, they interpret these data based on what "everyone" who is a member of the world of educational practice knows about what things mean and how things work in classrooms.[1]

That teachers do think and act in these ways to carry out student assessment is evident in the following test use project findings.

(1) In interviews, teachers report their uses of test results as serving most heavily the functions that are most central to teaching-as-practiced.

In the on-site interviews, teachers were able to describe with minimal constraints how they used test results and "data" from other assessment techniques. The purposes they most frequently cited were those that constitute their most essential work: deciding what to teach and how to teach it to students of different achievement levels; keeping track of how students are progressing and how they (the teachers) can appropriately adjust their teaching; and evaluating and grading students on their performance (See Table 1). Clearly, these are the day-to-day routines of teaching.

Less frequently, respondents mentioned using assessment results in deciding to refer students who need special instruction and to counsel, advise, and direct students. These are important teaching responsibilities, but ones that serve to support or facilitate more basic instructional work.

---

[1]These ways of describing what teachers do and think may sound a bit odd. If they do, it is because they come from a perspective that is not widely represented in the field of education or educational research: a branch or "school" of sociology known as ethnomethodology (e.g., Cicourel, 1974; Garfinkle, 1967; Mehan & Wood, 1975). Ethnomethodologists have studied how people do what they do in a variety of institutional settings; how juries make decisions (Garfinkle, 1967); how policemen on the beat decide that something seems amiss (Sudnow, 1972); how attendants in psychiatric wards decide how to handle patients (Wood, 1968); how educators place students in particular programs and classrooms (Kitsuse & Cicourel, 1963; Leiter, 1974); and so on. Ethnomethodogists' conceptualization of members of social groups as practical reasoners and decision makers is based on this kind of research. Thus, the analysis presented here--the view that teachers act as practical reasoners and decision makers as they go about evaluating students' performance--is not as unusual as the terminology makes it sound. In fact, it is an analysis grounded in a theoretical framewok derived from a substantial amount of research.

Table 1

Types of Tests and the Uses of Their Results (Interview Data, n=44)

USES Counts: Elementary Secondary Cell Total	Type of Test: Standardized	Curriculum-Embedded	District Objectives-Based	Minimum Competency	Statewide Assessment	School, Department, Grade-Level Teacher-Constructed	Diagnostic (Skills)	Other (informal)	TOTAL
Planning Instruction	9 4 13	8 2 10	3 0 3	1 3 4	2 0 2	1 2 3	11 13 24	1 1 2	13 8 21	49 33 82
Referral/Placement:	9 2 11	 0	 0	0 1 1	 0	0 2 2	2 1 3	 0	2 4 6	13 10 23
Within Classroom Grouping & Individual Placement	4 0 4	18 0 18	5 0 5	1 2 3	0 1 1	1 3 4	2 4 6	6 0 6	11 3 14	48 13 61
Holding Students Accountable for Work, Discipline	 0	3 0 3	 0	 0	 0	 0	4 4 8	 0	2 0 2	9 4 13
Assigning Grades	0 1 1	14 3 17	1 0 1	0 1 1	 0	0 5 5	15 17 32	1 0 1	7 1 8	38 28 66
Monitoring Students' Progress	 0	14 0 14	4 0 4	 0	 0	0 2 2	10 8 18	1 0 1	10 2 12	39 12 51
Counseling & Guiding Students	1 2 3	 0	2 0 2	 0	 0	 0	2 8 10	1 0 1	4 2 6	10 12 22
Informing Parents	 0	 0	1 0 1	 0	 0	 0	 0	 0	1 0 1	2 0 2
Reporting to District Officials, School Board, etc.	 0	1 0 1	2 0 2	 0	 0	 0	 0	 0	3 0 3	6 0 6
Comparing Groups of Students, Schools, etc.	1 0 1	 0	1 0 1	 0	 0	 0	 0	 0	1 0 1	3 0 3
Certifying Minimum Competence	 0	 0	 0	0 1 1	 0	 0	 0	 0	 0	0 1 1
TOTAL Use CITATIONS	24 9 33	58 5 63	19 0 19	2 8 10	2 1 3	2 14 16	46 55 101	10 1 11	54 20 74	217 113 330
Explicit Statements: "NOT USED"	5 5 10	 0	 0	1 1 2	0 7 7	1 0 1	 0	 0	0 1 1	7 14 21
Total Citations	29 14 43	58 5 63	19 0 19	3 9 12	2 8 10	3 14 17	46 55 101	10 1 11	54 21 75	224 127 351

Use of test results in such tasks as comparing groups of students and reporting to those at higher levels of the school and district organizational hierarchy were rarely mentioned. These matters are not in themselves unimportant. The reporting of scores to the school board, for instance, may be of considerable moment for the principal. Comparing classrooms or schools is often of central concern to district administrators and program coordinators. And these reports and comparisons may ultimately have an impact on teachers' daily professional lives. It is not that these activities are inherently trivial, then, that makes them non-salient for teachers; it is their remoteness from teachers' practical tasks that makes them so.

(2) The means of assessment on which most teachers rely most heavily are those which facilitate the accomplishment of their routine activities under the exigencies they face.

Reanalysis of data from an earlier CSE test use study (Yeh, 1978) found among 256 elementary school teachers surveyed that of all the tests they gave to their students, teacher-made tests figured more heavily than others in teachers' classroom decision making. The reanalysis also discovered that for assessing student progress teachers relied heavily on interactions with and observations of students.

On-site interviews supported and elaborated these findings. The 44 teachers interviewed collectively cited 351 uses for nine types of assessment techniques. (Refer again to Table 1.) They reported more uses (101) and more kinds of uses for their own, self-constructed tests and major assignments, e.g., essays, reports, etc., than for any other assessment type. Uses for other, less formal, teacher-developed strategies--peer evaluation, oral exercises, conferences with students, consultations with students' former teachers, etc.--were mentioned next most frequently (75 times) followed by curriculum-embedded tests available commercially or constructed by the

local school districts (63 times). Furthermore, in schools in each of the three districts studied, the aforementioned types of assessment techniques were those in which students spent the greatest proportions of their total assessment time.

National survey results dramatically confirmed the generality of these findings for both elementary and secondary teachers. Teachers were asked to rate information from various sources (tests and others) as crucial, important, somewhat important, unimportant, or not available for conducting four routine decision-making activities. For initially grouping or placing students in a curriculum, for changing students from one group or curriculum to another, and for assigning grades, nearly every survey respondent reported that "my own observations and students' classwork" was a crucial or important source of information (Refer to Tables 2 and 3). The great majority of respondents also indicated that the results of the tests they themselves developed also figured as crucial or important in these same decisions. Many elementary school teachers also responded that the "results of tests included with the curriculum being used" figured heavily in their planning of teaching and in placing and changing the placement of students. Far lower percentages of teachers rated the other types of information listed as crucial and important in carrying out any of these three activities.

Looking over all these findings, it is evident that the types of assesssment that most teachers rely most heavily on have three characteristics in common:

- Immediate accessibility; teachers can give them when they choose and see the results promptly
- Proximity between their intended purposes and teachers' practical activities

° Consonance from teachers' perspectives, between the content they cover and the content taught.

Each of these features responds to the exigencies of teachers' practical circumstances.

Teachers must accomplish their instructional work--initial planning, distributing students, teaching, continued planning, evaluating--within a temporal structure to which are attached normative expectations. Teaching units, marking periods, semesters, school years--these and other divisions of school time each have inherent points of closure. By those end-points, given amounts of learning are expected to be accomplished. Thus, time presses; teachers and their students must "progress;" decisions most often cannot wait (c.f. Jackson, 1968; Sarason, 1971; Smith & Geoffrey, 1968).

Not only is teaching time rapidly moving, it is also very full. Teachers interviewed during the exploratory field work were asked to detail the time they spent on various job-related activities in a normal school week. When their estimates were aggregated, elementary teachers' estimates averaged 357 hours a year spent outside the classroom, or about nine hours each week during the school year. High school teachers, on the average, seemed to be spending 600 hours a year or about 15 hours a week, on job-related tasks outside the classoom. And, of course, classroom time itself is constantly busy. Thus, teachers use means of assessment that are immediately accessible--that can be employed at the appropriate moment in the flow of on-going instruction, and for which results are quickly available.

Teachers also operate in an environment of accountability and concern. The decisions that they make matter, in varying degrees, to students' educational futures and life changes. Minimum competency laws, as well as court suits filed for "failure to educate," testify to the social pressures that bear upon teachers. That teachers recognize these pressures and strive to act with consonant concern and effort is evident (e.g., Lortie, 1975).

Thus, teachers use assessment techniques that they feel accurately measure what has been taught, that measure the effects of the instruction that they believe they have given. And in response to both time and accountability demands, as well as to their own concern with assessing accurately, they employ measures which match with the practical activities they must accomplish. In this regard, both the reanalysis and the field work found that teachers frequently use curriculum-embedded placement tests for placement and self-constructed and curriculum-embedded unit tests for tracking students' progress, for assessing performance on a unit, and for grading students. The exploratory on-site visits also discovered heavy use by instructional specialists (remedial reading teachers, teachers of the learning disabled, etc.), of normed diagnostic tests, e.g., the Sucher-Allred and the Bergantz Inventory of Basic Skills, for diagnosing individual learning problems and developing individualized programs.

In summary, the assessment techniques teachers seem to use most--teacher-made tests and assignments, curriculum-embedded tests, and especially the phenomenological data on students' performance that teachers gather daily in classrooms--respond to the practical exigencies teachers face and the routine tasks they must accomplish. In their use of these means of evaluating student achievement, teachers reveal themselves as practical reasoners and decision makers in their everyday professional lives.

(3) When test results are differentially important for teachers, their importance varies with their responsiveness to the practical exigencies that surround the task at hand.

As Tables 2 and 3 display, teachers rarely find standardized test results important in deciding on students' report card grades. However, substantially greater proportions of teachers report that they give standardized test results important consideration when it comes to planning their

Table 2

Elementary Teacher Use of Assessment Information for Different Decision-making Purposes

(Percentages reporting use of this information as crucial or important for the specified purpose)

Source/Kind of Information	Planning Teaching at Beginning of School Year		Initial Grouping or Placement of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Previous teachers' comments, reports, grades	57	52	62	55	x	x	x	x
Students' standardized test scores	57	54	57	52	55	53	17	16
Students' scores on district continuum or minimum competency tests	51	47	50	45	45	39	20	18
My previous teaching experience	94	94	x	x	x	x	x	x
Results of tests included with curriculum being used	x	x	78	67	83	82	75	77
Results of other special placement tests	x	x	61	56	x	x	x	x
Results of special tests developed or chosen by my school	x	x	x	x	56	52	42	42
Results of tests I make up	x	x	80	86	78	85	92	95
My own observations and students' classroom work	x	x	96	97	99	99	98	98

Table 3

High School Teacher Use of Assessment Information for Different Decision-making Purposes

(Percentages reporting use of this information as crucial or important for the specified purpose)

Source/Kind of Information	Planning Teaching at Beginning of School Year		Initial Grouping or Placement of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
	English	Math	English	Math	English	Math	English	Math
Previous teachers' comments, reports, grades	28	29	34	40	x	x	x	x
Students' standardized test scores	47	29	49	30	62	39	12	8
Students' scores on district continuum or minimum competency tests	48	30	47	36	53	36	9	5
My previous teaching experience	99	97	x	x	x	x	x	x
Results of tests included with curriculum being used	x	x	45	35	58	43	44	31
Results of other special placement tests	x	x	42	26	x	x	x	
Results of special tests developed or chosen by my school	x	x	x	x	50	31	28	34
Results of tests I make up	x	x	87	77	92	91	99	99
My own observations and students' classroom work	x	x	99	93	99	97	99	95

teaching at the beginning of the year. Standardized test scores also figure as crucial or important for many teachers as they go about the business of distributing and re-assigning students to instructional groups and curricula.

In the context of grading, standardized tests have qualities that are exactly the opposite of those assessment results that most teachers rely on most heavily. The classroom teachers interviewed, for instance, complained that standardized test scores for their current class(es) arrived in their hands too late in the school year to be of any use. In many cases, teachers never got them for this year's students: their results arrived the following fall. Many interviewees also noted that the scores provided little diagnostic information; others pointed out that the content of such tests overlapped only partially with what they were teaching. As usually scheduled and employed, then, standardized tests lack immediacy of accessibility. Their purposes are not perceived as proximal to teachers' everyday tasks (as one respondent put it, "they're for comparison, not diagnosis of my kids' weaknesses and strengths"). And many teachers perceive a poor fit between what they teach and what standardized tests cover.

Nevertheless, in the context of another activity, more teachers find standardized test results useful. At the beginning of the year, teachers can drop into the office and check the standardized test scores of their new class(es) as they plan what to teach and how to pace their teaching through the opening weeks of the semester. And where standardized scores are reported on the class rosters that teachers receive at the beginning of a new semester, some teachers interviewed said that they skimmed the scores, noted those student scores that deviated sharply from most students' scores on the list, then visited counselors to check on the placement of the students in

question. Thus, depending upon the context--i.e., on the activity at hand and the range of information available--the scores of a given type of test may or may not meet teachers' practical needs. In those contexts where they do, teachers take them into account. In those contexts where they do not, teachers generally disregard them.

The points made in the foregoing discussion add further detail to the portrait of the teacher as practical reasoner and decisionmaker.

Given the way the teachers' everyday world is organized, standardized tests are often impractical as sources of information. The scores they provide cannot be used in the work that constitutes day-to-day teaching -- tracking students' progress through units, adjusting instruction to fit ongoing achievement, assigning grades, etc. But, when practical circumstances allow and on those occasions where practical needs arise, teachers do treat standardized test results as important information. Thus, viewed from within "the world known in common and taken for granted" by teachers, teachers' demeanor toward and actions regarding standardized test scores make practical sense.

(4) For given activities and decisions, teachers most often use the results of various types of assessment techniques collectively. Scores from one test or one type of test rarely serve alone as the basis for accomplishing a task.

The on-site interviews indicated that teachers most often consider the results of several types of assessment techniques in carrying out a particular task. On the 351 instances in which teachers interviewed cited their uses for particular test scores and other assessment results, in 237 cases the scores and results were used as one of many information sources (See Table 4). Reanalysis of Yeh's (1978) research discovered the same phenomenon. In both pieces of research, which CSE used to plan test use project

Table 4

Overall Patterns of Assesment Results Use: Interview Data

Functional Importance

	Sole Source of Information Consulted	One of Several Major Sources	One of Many Sources	Verification Source	Not Used	Total
Instances Mentioned by 44 Teachers	18	65	237	10	21	351
	(5.1%)	(18.5%)	(67.5%)	(2.8%)	(6.0%)	(100%)

activities, it also became evident that teachers often revise decisions made on the basis of test scores in light of their ongoing experience with children in the classroom. Other research reports similar patterns of action by teachers (e.g., Airasian, 1979; Salmon-Cox, 1980; Shumsky & Mehan, 1974; Kitsuse & Cicourel, 1963; Leiter, 1974).

Once again, the results of the national survey substantiate these earlier project findings. This is indicated in the distribution of survey responses to those questions that ask teachers to report on the importance of different types of assessment information. (Refer to Tables 5 and 6.) Extremely high proportions of both elementary and secondary teachers' reported giving at least some importance to each type of information listed under three of the decision-making activities: initial planning, initial grouping and placement of students for instruction, and reassignment of students to different groupings and curricula. One need not examine the response patterns of individual teachers, then, to ascertain that the vast majority of them take a wide variety of kinds of assessment information into account in making each of these three types of instructional decisions. A glance at Table 7 shows more. Not only do survey respondents indicate that they consult several sources of information in students' achievement in making a particular instructional decision, they also report thinking that many kinds of assessment techniques give them crucial and/or important information.

Put another way, it does not seem as if teachers base their decisions primarily on one kind of assessment information, then look to others merely for confirmation or the sake of form. Rather, they appear to weigh various kinds of data on student achievement and to make sense of what the data mean more-or-less holistically. If this is in fact the case, it is a practice

Table 5

Proportion of Elementary Teacher Respondents Indicating Use of Information as "Somewhat Important," "Important," or "Crucial" for Each Task

Source/Kind of Information	Planning Teaching at Beginning of School Year	Initial Grouping of Students	Changing a Student from One Group or Curriculum to Another	Deciding on Students' Report Card Grades
Previous teachers' comments, reports, grades	93	75	x	x
Students' standardized test scores	92	91	89	43
Students' scores on district continuum or minimum competency tests	92	91	90	55
My previous teaching experience	100	x	x	x
Results of tests included with curriculum being used	x	98	97	93
Results of other special placement tests	x	96	x	x
Results of special tests developed or chosen by my school	x	x	96	81
Results of tests I make up	x	96	97	99
My own observations and students' classroom work	x	99	100	100

Table 6

Proportion of High School Teacher Respondents Indicating Use of Information as "Somewhat Important," "Important," or "Crucial" for Each Task

Source/Kind of Information	Planning Teaching at Beginning of School Year	Initial Grouping of Students	Changing a Student from One Group or Curriculum to Another	Deciding on Students' Report Card Grades
Previous teachers' comments, reports, grades	71	75	x	x
Students' standardized test scores	77	76	86	24
Students' scores on district continuum or minimum competency tests	78	78	83	26
My previous teaching experience	100	x	x	x
Results of tests included with curriculum being used	x	83	87	68
Results of other special placement tests	x	80	x	x
Results of special tests developed or chosen by my school	x	x	84	61
Results of tests I make up	x	97	98	100
My own observations and students' classroom work	x	99	100	99

Table 7

Proportion of Teachers who Report Considering Many Types of Assessment Information

Critical/Important for Given Activities

	Planning Teaching at Beginning of School Year	Initial Grouping or Placement of Students	Changing Grouping or Placement	Deciding on Report Card Grades
Number of Sources of Information Given in Question on Survey	4	7	6	6
Number of Sources Defined as "Many" for Purposes of this Analysis	3	4	4	4
Proportion of Elementary Teachers who Indicated That at Least this many functioned as Critical and/or Important for the Given Activity	50%	71%	62%	40%
Proportion of High School Teachers	33%	47%	49%	20%

typical of clinical professions. The sociologist Homans (1950) long ago pointed out:

> Clinical science is what a doctor uses at his patient's bedside. There, the doctor cannot afford to leave out of account anything in the patient's condition that he can see or test...It may be the clue to the complex...In action we must always be clinical. An analytical science is for understanding but not for action.

More recently Friedson (1970) has outlined other features of what he calls the "clinical mentality." He underscores that "the clinician is prone in time to trust his own accumulation of personal first-hand experience" and to be "particularistic," emphasizing the uniqueness of individual cases. This is evident in teachers' consistent reliance on the evidence of their personal, interactive experience with and observation of children in the classroom. It is also evident in many interviewees' remarks about why the results of one test or one type of test--or even tests in general-- cannot be trusted without reference to everyday experiential evidence.

> I don't rely heavily on a lot of the test scores because I find that...some students are test takers and other are not...some students can handle the format, the time limit, (but in many cases) students are capable of more than test scores show.

> I hate to say it, but I'd say about a third of these students don't give it their best shot. They feel there's nothing in it for them. There's no grade for it; there's no use for it--so they don't care.

> If I see there are certain kids having trouble I may look at their folders and find out about them. But I try not to be swayed by somebody else's judgement...I may get more out of them by what I'm telling and trying to motivate them to do better than they've ever done before.

> You can't count a score on one test too heavily. The kid could be sick or tired or just not feel up to doing it that day. Maybe his parents had a fight the night before. Maybe he doesn't try. Maybe he doesn't test well.

Numbers of other respondents voiced equivalent opinions.

Similar findings appeared when teachers' opinions of the factors which can influence test scores were elicited in a closed-ended format in Yeh's (1978) questionnaire study. On a five-point rating scale (where 5 = "great influence" on test scores), among the factors for which teachers rated influence as 3.0 or higher were the following: Students' test-taking skills ($\bar{X}$=4.4); directions, content, format, physical characteristics, student motivation ($\bar{X}$=4.3); unusual circumstances--special activities, distractions ($\bar{X}$=4.2); and parent interest ($\bar{X}$=3.0).

Part of what "everyone knows" in the world of educational practice, then, is that students vary as test takers and that a variety of situational factors can influence students' test performances. Better, then, to rely on a variety of sources of information -- especially one's day-to-day, first hand observations of and interactions with the individual across a variety of recurrent performance settings in the classroom -- and to make sense of all the data at hand "in this situation" in light of one's practical knowledge, one's clinical experience.[2]

(5) Teachers' explicit comments on tests and testing orient to the routine constitutive tasks and exigencies of teaching-as-practiced.

The above evidence warranting the concept of the teacher as practical reasoner and decision maker is based on what teachers say that they do in using tests. Another slightly different form of evidence--what teachers

[2]Perhaps the data and analysis presented here explain why an overwhelming percentage of survey respondents teaching at both the elementary and secondary levels agree that minimum competency tests should be required of all students for promotion at certain grade levels or for high school graduation, while simultaneously agreeing that teachers should not be held accountable for students' scores on minimum competency or standardized achievement tests. See Tables 8 and 9.

report that they believe and think--ratifies the same concept. In fieldwork interviews, teachers' remarks repeatedly called attention to their need for tests that are immediately accessible, that are consonant with the material taught, and that produce results that can function in the routine tasks they confront everyday. The following quotations are illustrative of these points.

> The ITBS is almost useless in the spring, which is too bad, because I feel there is some valuable information there, progress and growth. But we get the scores the last week of school.
>
> That computer-processed data (on district, objectives-based tests) can really be used with those kids that need help. It does a better job of identifying students and student needs...I can now say 'the kid needs to work on objectives 2, 3, 5, and 9.'
>
> I don't feel we need to test, test, test--but if the information is something I can use to prescribe instruction, then I don't really mind giving it.
>
> In math, you know, it's a good idea to keep them (tests) in my class. As long as testing stays in math class it seems like it fits in, 'cause tests are part of taking math.
>
> In my class, I like to use the criterion-referenced test of basic skills. The tests are geared to certain basic skills the book's developing -- vocabulary, spelling, and writing.
>
> The district (testing) design is important because it's the only thing you can pass on to other schools which is meaningful to everybody.
>
> I don't use (the results of the reading series tests) unless there are results that completely throw me--like someone who usually does a good job completely bombed one--then I'll do something about that, try to find some extra work to go over it.

The orientation to assessment "for all practical purposes" that emerges in these fieldwork interview remarks appears again in the reanalysis of

Yeh's (1978) data. There, on a five-point rating scale where 5 = "Very Important," teachers rated the following considerations for selecting tests as high: test material is similar to what I present in class ($\bar{X}$=4.5); the test has clear format, pictures, directions ($\bar{X}$=4.6); the test accurately predicts student achievement ($\bar{X}$=4.4); the test is simple to administer and/or score ($\bar{X}$=4.2). These practical matters in test selection are consonant with the patterns of teachers' concerns and actions reported throughout this section.

SUMMARY

A variety of routine tasks constitutes the world of teaching-as-practiced. Teachers must accomplish these in a context characterized by recurrent time limits, others' demands for high performance and accountability at those deadlines, and their own concerns with providing effective and appropriate instruction. These features of the world of teaching-as-practiced impinge upon teachers' testing practices and test use. Their reasoning and decision making about assessment and its uses are structured by and oriented to their practical circumstances.

The purposes for which they use assessment results most often are those inherent in the most central activities of teaching as it is practiced: determining what to teach and how to teach it in general and to various class members in particular, determining from day to day whether it is being learned and adjusting instruction as necessary to be sure it is; and giving students grades so that they and their parents will know how they are doing. For those purposes less intimately connected with the central work of teaching, use of assessment results seems to occur less frequently. Action, in the "world known in common and taken for granted" by teachers, centers on the work of daily instruction.

The tests teachers use most frequently are those that fit their practical circumstances: formal and informal measures they themselves construct or seek out for the information they provide; curriculum-embedded tests that come with commercial or district materials. These are immediately accessible, proximate in purpose to the tasks teachers must accomplish, and consonant with the material taught. The further that tests and testing features are removed from these qualities, the less likely their results seem to be used.

The way in which teachers use tests follows from their practical understandings of the "scenic features" of their world. They recognize--tacitly in their actions and often explicitly in their words--that performance varies with context and that many "readings" of student achievement are better than few. Thus, they most often use results from many assessment types collectively to accomplish given purposes. Their immediate, recurring experience with children often over-rides scores from paper-and-pencil instruments.

Teachers' comments about tests and testing confirm their orientation to the practical business of getting everyday tasks done in time and done well. They speak of the need to diagnose, prescribe, and assess efficiently and accurately. They talk of the need for test directions and formats that are clear. And they comment practically about the need to consider "extenuating circumstances," to pass on information "which is meaningful to everybody," and the like.

It should be apparent in all that I have said up to now that teachers' attitudes toward the assessment of student achievement in general -- and toward testing in particular -- are neither universally negative or globally positive. Attitude questions on the national survey confirm that this is

the case and, once again, reflect teachers' practical concerns. (See Tables 8 and 9.) My intent here is not to examine teachers' responses to these questions in detail, but merely to point out that they tend to support the analysis presented through the preceding pages.[3] Thus, for instance, most teachers see testing as a technique that motivates students to study harder (elementary = 73 percent; high school English = 80 percent; high school math = 93 percent). Perhaps with this in mind, most teachers also agree that tests of minimum competency should be required of all students for promotion or graduation. (See item #10 in Tables 8 and 9.) Yet, at the same time, there is substantial concern that minimum competency tests "are frequently unfair to particular students" (elementary teachers agreeing = 58 percent; high school English and math teachers, 48 percent; 35 percent). Moreover, many teachers also worry that minimum competency testing affects "the amount of time I can spend teaching subjects or skills that the tests do not cover" (elementary = 2 percent; high school English and math, 62 percent; 42 percent). These responses clearly reflect teachers' practical orientation toward testing: their concerns with motivating students, with the student as an individual, and with the effect of testing on their discretion as experienced clinicians to decide what is appropriate to teach to their practical students.

A little over 60 percent of the teachers feel that the tests developed in their districts are very good. Most elementary teachers (59 percent) and many high schools teachers (46 percent in both subject areas) find that the

---

[3]Teachers were asked to indicate their attitudes on a four-point scale where 4 = strongly agree, 3 = agree, 2 = disagree, 1 = strongly disagree. The tables show the proportion of teachers who chose either of the first two categories.

Table 8

Elementary Teacher Attitude Toward Tests and Test-Related Issues
(N=486)

	Item	Percentage of Teachers in Agreement
(1)	Testing motivates my students to study harder.	73
(2)	Commercial tests are usually of high quality.	59
(3)	The content (or skills) on most required tests is very similar to the content (or skills) that I teach.	77
(4)	The pressure that testing exerts on the schools has a generally beneficial effect.	48
(5)	Recently, I have been spending more teaching time preparing my students to take required tests.	46
(6)	The tests developed in our district are very good.	62
(7)	The curriculum today demands more complex student thinking than in the past.	74
(8)	Teachers should not be held accountable for students' scores on standardized achievement tests or tests of minimum competency.	71
(9)	In our school, students are more rigidly tracked than they were two or three years ago.	58
(10)	Tests of minimum competency/proficiency/ functional literacy should be required of all students for promotion at certain grade levels or for high school graduation.	81

Table 8
(continued)

Item	Percentage of Teachers in Agreement
(11) Tests of minimum competency are frequently unfair to particular students.	58
(12) As a result of minimum competency tests (and similar programs), parents are contacting schools about their greater numbers.	53
(13) Tests of minimum competency have affected (would affect) the amount of time I can spend teaching subjects or skills that the tests do not cover.	62
(14) In our school, testing programs are generally held to be much less important than the social problems with which we are concerned.	39
(15) Basic skills teaching (including remedial work) is now consuming a substantially increased proportion of our school's educational resources.	88
(16) The proportion of our schools resources now allocated to basic skills teaching is so great as to detract from the quality of out total educational program.	23

Table 9

High School Teacher Attitude Toward Tests and Test-Related Issues
(N=365)

	Item	Percentage of Teachers in Agreement: English	Math
(1)	Testing motivates my students to study harder.	80	93
(2)	Commercial tests are usually of high quality.	46	46
(3)	The content (or skills) on most required tests is very similar to the content (or skills) that I teach.	77	79
(4)	The pressure that testing experts on the schools has a generally beneficial effect.	60	72
(5)	Recently, I have been spending more teaching time preparing my students to take required tests.	41	30
(6)	The tests developed in our district are very good.	62	60
(7)	The curriculum today demands more complex student thinking than in the past.	62	54
(8)	Teachers should not be held accountable for students' scores on standardized achievement tests or tests of minimum competency.	61	61
(9)	In our school, students are more rigidly tracked than they were two or three years ago.	42	36
(10)	Tests of minimum competency/proficiency/functional literacy should be required of all students for promotion at certain grade levels or for high school graduation.	86	90

Table 9
(continued)

Item		Percentage of Teachers in Agreement: English	Math
(11)	Tests of minimum competency are frequently unfair to particular students.	48	35
(12)	As a result of minimum competency tests (and similar programs), parents are contacting schools about their children more frequently or in greater numbers.	42	36
(13)	Tests of minimum competency have affected (would affect) the amount of time I can spend teaching subjects or skills that the tests do not cover.	62	42
(14)	In our school, testing programs are generally held to be much less important than the social problems with which we are concerned.	32	42
(15)	Basic skills teaching (including remedial work) is not consuming a substantially increased proportion of our school's educational resources.	84	74
(16)	The proportion of school's resources now allocated to basic skills teaching is so great as to detract from the quality of our total educational program.	28	21

quality of commerical tests is usually high. And over three-quarters of the teacher respondents observe that the content or skills on required tests are very similar to what they teach (see item #3 in Tables 8 and 9). In short, teachers are certainly not "anti-testing" in any general sense, as some studies have concluded. They are simply concerned that test information serve them: (1) as they go about doing the daily work that is at the core of teaching as practiced, and (2) that testing serve them efficiently in the context of the practical contingencies and exigencies that they face.

SOME IMPLICATIONS FOR LOCAL POLICY AND PRACTICE

I suggested at the outset of this paper that if testing programs are to be useful to teachers and used in classrooms, they must take into account teachers' routine thinking and practices in assessing students' achievement. To review, such programs would feature tests that are

(1) proximal to the everyday instructional tasks teachers need to accomplish - planning their teaching, diagnosing students' learning needs, monitoring their progress through the curriculum-as-taught, placing students in appropriate groupings and instructional programs, adjusting their teaching in light of students' progress, and informing parents and others how students are doing;

(2) consonant, from teachers' perspectives, with the curriculum that teachers are actally teaching;

(3) immediate accessible to teachers, so that teachers can give them to students when the time seems appropriate and have the results available promptly;

(4) designed to include a variety of performance "contexts," i.e., different types of response formats and tasks.

Many districts' (and schools') testing programs fail to meet these criteria in one or more ways. When they do, they become simply an extra burden for teachers. Instructional time is taken up in testing, but there are few

concomitant benefits for teachers or students. In other cases, districts (and sometimes schools) hope to meet the above criteria by developing sets of tests oriented to local curricular objectives. But the test use project's earlier interviews and continuing fieldwork indicate that in many cases these objectives-based tests only seem to meet the criteria listed above. Thus, the experience of one district studied by the project may provide a useful example of how those criteria can be met.

A Case in Point

The (mid-western) district in question (enrollment about 5,000) did not have vast resources. Nevertheless, it involved teachers during the school year and especially during the summer in building curricula and tests to accompany them. Teachers were participants in substantial numbers. (And at the elementary level, they were the leaders of cross-grade-level teaching teams -- leaders chosen by their colleagues.)

The emphasis in these recurrent projects was upon curricular objectives and instructional materials. An effort was made to select objectives and design materials that teachers found appealing and used. Repeated revisions of instructional materials and goals based on teachers' criticisms were part of the process. Tests were designed to fit each curriculum -- tests that met the teachers' routine teaching needs. Thus, the curricular packages included placement tests, chapter and unit tests, and semester and end-of-the-year review tests or "finals." These tests were also revised in response to teachers' criticisms during the development process, which included as a final step using the curricula and tests in schools throughout the district on a pilot basis for a year.

The tests themselves were designed to be computer scored and analyzed, using computers that the district had originally purchased for computer-

assisted-instruction in the high school. Teachers gave the tests at times that they felt were appropriate, turned them in for scoring, and received the analyzed results within a day or two. The results themselves came in the form of a set of sheets, one for each student. The sheet listed (1) each objective the test covered, (2) the number of items that assessed performance on each objective, (3) the number of items that the student passed and missed on each objective. At the top of the sheet was a paragraph listing the main types of errors that the student had made and stating just what problems the student seemed to be having. This was based on an analysis of the questions missed and the incorrect items chosen.

Teachers reported that they and their colleagues routinely used these tests. And interview response patterns indicated that they spent less time designing, administering, and scoring their own tests than teachers in the other districts visited. Interviewees stated explicitly that they used these tests (1) because they fit well with what they were actually teaching, (2) because they could be used flexibly, e.g., at any time, with one child or an entire class, (3) because scores came back promptly, (4) and because the analyses summarized information in a way that gave them precise diagnoses they could act on in placing students, in deciding who needed additional help on what skills, etc. In fact, the only complaint teachers made was that all the tests were multiple-choice tests. As one teacher put it, "that's a problem, 'cause sometimes you wonder whether they can apply the skills or ideas another way."

In short, this district made considerable efforts to assure that its testing program was useful to and used by teachers. In so doing, its program for testing fulfilled three of the four criteria identified earlier. The program met district needs, too. Semester and end-of-the-year finals

functioned to indicate the strengths and weaknesses of the students in particular schools and in schools throughout the district from year to year. Thus, they served various evaluation and managment functions.

Testing programs which take into account teachers' routine thinking and practices in assessing students' achievement can probably take many shapes. This is only one example. But it should be clear that programs of testing that ignore how teachers think and act toward student assessment can result in inefficiency and teacher resentment.

REFERENCES

Airasian, P. W. The effects of standardized testing and test information on teachers' perceptions and practices. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979.

Airasian, P. W., Kelleghan, T., Madaus, G. F., & Pedulla, J. J. Proportion and direction of teacher rating changes of pupils' progress attributable to standardized test information. Journal of Educational Psychology, 1977, 69(6), 702-709.

Berman, P., & McLaughlin, M. W. Federal programs supporting educational change, Vol. VIII: Implementing and sustaining innovations. R-1589/8-HEW. Santa Monica, CA: Rand, 1978.

Cicourel, A. V. Cognitive sociology. New York: The Free Press, 1974.

Dawson, J. E. Why do demonstration projects? Anthropology and Education Quarterly, 1977, 8(2), 95-105.

Deal, T. E. Linkage and information use in educational organizations. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, 1979.

Friedson, E. The profession of medicine: A study of the sociology of applied knowledge. New York: Dodd, Mead, 1970.

Garfinkle, H. Studies in ethnomethodology. Englewood Cliffs, New Jersey: Prentice-Hall, 1967.

Goslin, D. A. Teachers and testing. New York: The Russell Sage Foundation, 1967.

Homans, G. The human group. New York: Harcourt, Brace and World, 1950.

Jackson, P. Life in classrooms. New York: Holt, Rinehart & Winston, 1968.

Kitsuse, J. I., & Cicourel, A. V. The educational decisionmakers. Indianapolis, Indiana: Bobbs-Merrill, 1963.

Leiter, K. C. W. Ad hocing in the schools: A study of placement practices in two kindergartens. In A. V. Cicourel, et al., Language use and school performance. New York: Academic Press, 1974.

Lortie, D. C. Schoolteacher: A sociological study. Chicago: The University of Chicago Press, 1975.

Mehan, H., & Wood, H. The reality of ethnomethodology. New York: Wiley Interscience, 1975.

Meyer, J. W., & Rowan, B. The structure of educational organizations. In M. W. Meyer and Associates (Ed.), Environment and organizations. San Francisco: Jossey-Bass, 1978, 78-107.

Montjoy, R. S., & O'Toole, Jr., L. F. Toward a theory of policy implementation. Public Administration Review, 1979, 39(5).

Resnick, L. B. Introduction: Research to inform a debate. Phi Delta Kappan, 1981, 62(9), 623-625.

Sarason, S. The culture of the school and the problem of change. Boston: Allyn & Bacon, 1971.

Smith, L., & Geoffrey, W. The complexities of an urban classroom: An analysis toward a general theory of teaching. New York: Holt, Rinehart, & Winston, 1968.

Stetz, F., & Beck, M. Comments from the classroom: Teachers' and students' opinions of achievement tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, 1979.

Sudnow, D. (Ed.) Studies in social interaction. New York: The Free Press, 1972.

Weatherly, R., & Lipsky, M. Street-level bureaucrats and institutional innovation. Harvard Educational Review, 1977, 47(2).

USING TESTS:
WHO DO WE BELIEVE AND WHAT DOES IT MEAN?

James Burry

INTRODUCTION

In the first section of this paper I provide a review of the few previous studies of test use--what they say about use or non-use of certain kinds of test information and the explanations they offer in support of their conclusions.

Next I present the findings from CSE's test use survey showing teachers' stated uses of assessment information for specific classroom decisions. This section begins to develop some alternative reasons for why teachers value, or do not value, certain kinds of test information.

I discuss these reasons under the heading of school/district characteristics bearing on test use. These characteristics reflect the tests/testing resources provided to teachers, the kinds of assistance in testing activities they receive from their school or district, the "messages" they get about district testing policy on the basis of district uses of test data, and how quickly they get back test results from the district and whether they are in a format useful for instructional purposes. This section concludes with an interim summary and suggests alternative testing practices on the basis of our survey data and the fieldwork which preceded it.

In the final section of the paper, I draw implications and recommendations from the data, and suggest some methodological, technical, and organizational considerations that will need to be addressed before improvements in testing practice can begin.

PREVIOUS STUDIES OF TEST USE

The relatively few studies of teacher uses of tests have focused almost exclusively on standardized tests. These studies have described the uses, or non-uses, of standardized tests by teachers and some have gone on to explore some of the reasons for non-use.

Uses of Standardized Tests Ascribed to Teachers

Goslin reported in 1967 that elementary school teachers use standardized test results primarily to diagnose individual difficulties and to provide feedback to the student. However, he also reported that the teachers did not rely heavily on this kind of information. Less than 20 percent of the teachers had altered a course, and less than one third reported changing their methods as a result of standardized tests (Goslin, 1967).

Stetz and Beck (1979), in conjunction with the standardization of the Metropolitan Achievement Tests, conducted a study of teacher's opinions of the use and usefulness of standardized tests. Teachers in this study frequently responded that they used standardized test results for diagnosing strengths and weaknesses, measuring student growth, and

evaluating individual students. The finding that 80 percent of the teachers reported making only some or little use of the data from standardized tests is similar to conclusions reached by Goslin.

The Royal Oak Study (Boyd, et al., 1975) suggested that teachers do not rely on the results of standardized tests for decision making. Although teachers in this study reported variable use of results from the district-mandated testing program, there was little evidence that the testing program influences school curriculum or classroom instruction.

A study of standardized tests was recently reported by a group of researchers at the University of Pittsburgh and Carnegie-Mellon University (see Kappan, May 1981, pp. 623-636, for the five articles dealing with this study). The study was conducted in 18 school systems in western Pennsylvania. Data from the study came from 58 administrators and 68 teachers.

In the first of these articles, Resnick (1981) reports that school administrators and teachers rely more on direct observation and conversation with confidants than on information from standardized tests. In one of the companion articles, Sproull and Zubrow (1981) discuss the interviews they conducted with 58 administrators--none of whom were building-level administrators--and report that testing does not enjoy a very high status in most school systems. The study goes on to suggest that administrators think standardized tests are used for individual diagnoses and placement, instructional program evaluation, end-of-year achievement measurement, and reporting to outside agencies, and that they also believe that the benefits of testing accrue primarily to teachers and principals.

One of the other articles in this series (Salmon-Cox, 1981) discusses the results of interviews conducted with 68 elementary teachers on their uses of standardized tests. The teachers in this study most frequently mentioned observation as their favored assessment technique and, when they did refer to standardized tests, use consisted of supplementing other information, guiding instruction, and grouping and tracking students. However, when asked who would care if standardized tests were abolished, 45 percent of the interviewees replied that teachers would care, because teachers like to have a variety of information sources about children.

Reasons for Non-use of Standardized Tests

According to the Royal Oak study previously cited (Boyd, et al., 1975) teachers felt, for the most part, that standardized tests were selected by administrators and imposed on teachers, and did not furnish them with any new information to begin with. Although some teachers thought the test results were useful, most felt that the tests given were not useful for plaining instruction.

Based on the responses of the teachers in their study, Sproull and Zubrow (1981) reason that standardized tests measure only cognitive goals and not the social goals which their teachers stressed, and that while such tests partially measure a child's achievement they are not the broadly-based tests that teachers seem to prefer. On the other hand, Sproull and Zubrow assert that teachers also fault standardized tests because they are neither sufficiently precise for diagnostic purposes nor are they linked to instruction.

Other studies have criticized standardized tests for their inefficiency, narrowness of foci, breadth of foci, bias, invalidity, and unreliability (Broekhoff, 1978; Howe, 1978; Klein, 1970; Perrone, 1978; Burry, 1981a). Still others have dealt with the effects of testing on teachers' perceptions and practices (Airasian, 1979; Airasian, Kelleghan, Madaus, & Pedulla, 1977).

Teachers' lack of training is sometimes cited as bearing upon test use. Goslin (1967) found that less than 40 percent of all teachers have had minimal formal training (one course) in test and measurement techniques; that teachers, however, tend to view standardized tests as relatively accurate measures of student achievement, and see the abilities measured by these tests as important determinants of academic success; but that teachers make only limited use of these tests in grading and advising pupils and in providing them with feedback.

Hastings, Runkel, and Damrin (1961) also believe that test use depends on teacher knowledge of tests and how to interpret them. This belief is supported by a number of texts (e.g., Gorow, 1966) offering teachers information on building their own tests and improving them through analysis of test results. It is also seen in work like Bauerfeind's (1963) dealing with validity and reliability and designing a good testing program. Ebel (1967) called for inservice workshops to provide teachers with training in tests and testing issues. There is little evidence in our study that this call has been heeded.

The Question of Focus

Although most of the studies discussed here purportedly deal with

standardized, norm-referenced tests, some do not always keep their focus to the forefront; in secondary discussion of or allusion to primary work on standardized testing, the focus is equally subject to drift. For example, a treatment of standardized test use will frequently lose that qualified referent and begin to discuss "tests" or "testing" as though these phenomena had a uniform mode of expression. A work might begin with a discussion supposedly limited to standardized, norm-referenced tests--which are one particular kind of achievement test, loosen the focus with references to "tests," switch the focus again with references to "achievement tests," "ability test," and so forth. In this way, conclusions drawn about use or non-use of standardized, norm-referenced tests are on the one hand weakened since the focus shifts, but on the other hand are given unwarranted interpretive breadth when statements (critical of favorable) supposedly about standardized tests are framed in such a way that they may be taken as statements about achievement tests in general.

The range of reported or perceived uses of standardized tests is catholic: diagnosing individual student strengths and weaknesses; measuring student growth; end-of-year achievement measurement; instructional program evaluation; guiding instruction; grouping and tracking students; reporting to outside agencies, and so on. This seeming ubiquity is reflected in the criticisms of standardized tests: breadth of focus; narrowness of focus; cognitive focus; external focus. Viewing the test use literature as a body, the feeling conveyed is that it is legitimate to criticize any single test--standardized or other--because it cannot accomplish conflicting purposes nor embody competing properties.

Another impression sometimes conveyed is that users of tests, such as teachers, are concerned with discrete decisions and make those decisions, or would like to make them, on the basis of a single source of information, such as a formal act of testing. Some of the more recent work (Airasian, et al., 1977; Arasian, 1979; Salmon-Cox, 1981) does not evoke this picture of tests and decision making. For example, Salmon-Cox correctly stresses that teachers tend to rely on a variety of information as they make decisions about their students. Since teachers often refer to multiple sources of information to make a series of related instructional decisions, then if they perceive the purpose of an investigation is to ask teachers to describe the value of any single test--again, standardized or other--for any discrete decision, they will very likely find that test wanting.

A useful point of departure in some recent work (e.g., Bank, Williams, & Burry, 1981) suggests that standardized, norm-referenced tests can be faulted because they do not provide diagnostic and prescriptive linkages between testing and instruction.

CSE's test use work, which addresses these linkages, has sought to discover, directly from teachers, what kinds of information they rely on as they make their classroom decisions. In this context, our work did not focus only on standardized testing; rather it focused on those assessment activities--test and non-test, norm-referenced and criterion-referenced, formal and informal--that teachers use, frequently in some combination, to make decisions about individuals, groups, and classes. With this focus, as we shall see, teachers provide a somewhat different view of test use, whether standardized or in some other form. Our work, therefore, fills in

some of the gaps in our knowledge of uses of standardized tests, as well as of teacher-made assessments, curriculum embedded tests, school or district constructed tests, observation, and so forth. In this context, teacher statements about any single source of information assume less of an adversarial posture, and rather reflect the relative weights teachers assign to a range of assessment techniques set against a range of legitimate information needs.

TEACHERS' USE OF TEST AND OTHER INFORMATION: THEIR RELATIVE IMPORTANCE

This section provides a summary of our teachers' descriptions of the importance they place on various kinds of information for specific decision-making purposes. These decision areas are: (1) planning teaching at the beginning of the school year; (2) initial grouping or placement of students for instruction; (3) making decisions to change a student from one group or curriculum to another, or to provide remedial or accelerated instruction; and (4) making decisions on students' report grades.

Before I discuss our teachers' responses, let me offer a point or two about how they seem to feel about tests in general; in one or two respects their attitudinal statements differ from attitudes ascribed to teachers in earlier research.

About 80 percent of all our teachers--elementary and secondary--described the content of their required tests as being similar to what they teach. Lest this be seen as an implication that required testing is having

a levelling effect on the curriculum, the same percentage also agreed that the proportion of school resources allocated to basic skills teaching is not high enough to detract from the quality of the school's total educational program.

About 75 percent of the elementary teachers and 85 percent of the secondary teachers feel that testing motivates their students to study harder--which surely has instructional implications.

As a final attitudinal example, about one-half of the elementary teachers and about two-thirds of the secondary teachers stated that testing exerts a generally beneficial effect on their schools.

I'll now talk about the test use responses from the elementary teachers, then the secondary teachers. The data appearing in Table 1 following indicate the percentage of elementary teachers, broken down for reading and math, who rated a variety information sources as crucial or important for making the decisions of interest. Numbers in parenthesis reflect percentages of teachers reporting that the assessment information is not available.

The Elementary Teacher

Several conclusions are suggested by these data. For example, whether a respondent is describing assessment information use for reading or math, the relative weight elementary teachers ascribe to a given kind of information remains fairly constant in the decision-making process.

In planning for instruction, the individual teacher's previous classroom experience is by far the single most important kind of information. Students' scores on standardized tests and on district continua or minimum

Table 1

Elementary Teacher Use of Assessment Information for Different Decision-Making Purposes

Percentages reporting use of this information for the specified purpose.
Numbers in parenthesis reflect percentages of teachers reporting the information source as not available.

Source/Kind of Information	Planning Teaching at Beginning of School Year		Initial Grouping of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Previous teacher's comments, reports, grades	57 (1)	52 --	62 --	55 --	--	--	--	--
Students' standardized test scores	57 --	54 (1)	57 (1)	52 (2)	55 (1)	53 (2)	17 (7)	16 (7)
Students' scores on district continuum or minimum competency tests	51 (17)	47 (19)	50 (20)	45 (22)	45 (20)	39 (24)	20 (22)	18 (23)
My previous teaching experience	94	94	--	--	--	--	--	--
Results of tests included with curriculum being used	--	--	78 (6)	67 (10)	83 (2)	82 (4)	75 (3)	77 (3)
Results of other special placement tests	--	--	61	56	--	--	--	--
Results of special tests developed or chosen by my school	-- --	-- --	-- --	-- --	56 (21)	52 (23)	42 (23)	42 (24)
Results of tests I make up	--	--	80 (6)	86 (3)	78 (5)	85 (1)	92 (2)	95 (1)
My own observations and students' classroom work	--	--	96	97	99	99	98	98

competency tests, however, appear to be as important in this decision as comments and other information about students offered by their previous teachers. Note that for about 20 percent of these teachers no district or minimum competency test (MCT) data are available at the beginning of the school year.

In making their initial grouping decisions, the elementary teachers' own observations and their own tests are deemed most important by most teachers, followed by curriculum-embedded tests, other special placement tests, and previous teacher comments. Again, about 20 percent of the teachers state that no district continua or MCT data are available.

For a sizeable number of teachers, more than 50 percent of the sample, students' scores on standardized tests are also important for initial placement decisions. But note that these tests are also important for decisions about changing a student from one group to another or one curriculum to another. That is, for a sizeable number of elementary school teachers, standardized test scores assume importance not only at the beginning of the school year but also during the school year.

With regard to the elementary teachers' decisions about changing a student from one group or curriculum to another, teacher observation is still most important for most teachers. In this decision area, however, most teachers seem to place almost equal weight on their own tests and curriculum embedded tests. This group of tests appears second, then, in order of importance, followed by the results of special school tests and standardized tests which appear roughly equal in value, and district continua or MCTs which are deemed useful by the smallest percentage of

teachers. Note, however, that the latter are not available for better than 20 percent of the teachers, and that similar percentages report non-availability of special school tests.

A similar weighting pattern appears for decisions about students' report card grades, with the exception that here the percentages of teachers ascribing importance to student scores on standardized and district continuum or competency tests fall off quite markedly, and drop to a somewhat lesser degree in the case of special school tests. Patterns of test non-availability also remain constant.

Elementary teachers appear, then, to rely on multiple sources of information for making their classroom decisions. Use of the more "formal" tests is more prevalent early in the school year, and as the year advances and different kinds of decisions about individual students, groups, and classes have to be made, teachers seem to switch more to use of their own professional experience, observations, students' classroom work, the results of teacher-made tests, and tests that come with the curriculum informing their teaching. This does not mean that any single measure entirely dominates or drops from the decision process.

The Secondary Teacher

I turn now to the secondary teachers' response to the same questions of test use. Table 2 following shows the percentages of secondary teachers, with separate entries for English and math teachers, who rated a given information source as crucial or important for the specified decision concerns. Numbers in parentheses indicate percentages of teachers for whom the assessment information is not available.

Table 2

Secondary Teacher Use of Assessment Information for Different Decision-Making Purposes

Percentages reporting use of this information for the specified purpose.
Numbers in parenthesis reflect percentages of teachers reporting the information source as not available.

Source/Kind of Information	Planning Teaching at Beginning of School Year		Initial Grouping of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
	English	Math	English	Math	English	Math	English	Math
Previous teacher's comments, reports, grades	25 (9)	31 (9)	34 (10)	40 (11)	--	--	--	--
Students' standardized test scores	47 (3)	28 (4)	46 (5)	33 (16)	64 (1)	36 (10)	12 (14)	9 (24)
Students' scores on district continuum or minimum competency tests	47 (18)	26 (27)	47 (19)	36 (30)	55 (16)	37 (31)	8 (26)	4 (37)
My previous teaching experience	99	96	--	--	--	--	--	--
Results of tests included with curriculum being used	--	--	41 (25)	34 (38)	59 (12)	46 (25)	44 (14)	34 (27)
Results of other special placement tests	--	--	39 (26)	26 (37)	--	--	--	--
Results of special tests developed or chosen by my school	-- --	-- --	-- --	-- --	49 (25)	32 (46)	27 (26)	23 (40)
Results of tests I make up	--	--	85 --	76 (5)	90 --	90 (4)	100 --	96 --
My own observations and students' classroom work	--	--	98	90	99	97	99	95

As was the case with elementary school teachers, the secondary English and math teachers' previous experience is by far the most important source of information as they plan instruction at the beginning of the school year. For the English teachers, students' scores on standardized tests and their scores on district continua or tests of minimum competency are held as important by almost half of the sample, followed by previous teachers' comments with about 25 percent. For the math teachers, only about 25 percent report importance for standardized and district continuum tests. Note that for students' scores on district continua/minimum competency tests, almost 20 percent of the English teachers and almost 30 percent of the math teachers report this kind of assessment information is not available to them.

In making their decisions about initial grouping or placement of students, secondary teachers' own observations and the results of tests they make up themselves are deemed most important, with the results of standardized tests, district continua and MCT, and curriculum-embedded tests roughly equal and next in order of importance. Previous teachers' comments are about the same for English and math teachers; 34 percent of the English teachers and 40 percent of the math teachers report these sources as important in this decision area; higher percentages of English teachers place importance on other special placement tests than do math teachers.

Again, as was the case with the elementary teachers, note that students' scores on formal tests continue to have importance for a sizeable number of secondary teachers as they make their initial grouping decisions;

this trend is somewhat more pronounced for the English teachers, with almost half them reporting those tests as important, but with only 30-odd percent of the math teachers agreeing. Note once again that for a sizeable number of teachers, certain kinds of test information are reported as not available: about 20 percent of the English teachers and 30 percent of the math teachers report there are no district continua/minimum competency test data; anywhere from 25 to almost 40 percent of the secondary teachers state there are no tests available as part of their curricula and no special placement information.

In terms of secondary teachers' decisions about changing a student from one group or curriculum to another, teachers' observations and results of their own tests are the most important sources of information for most teachers. For the English teachers, the next most important kinds of information, in descending order, are standardized tests, curriculum-embedded tests, district continua or MCT, and special school tests. For the math teachers, the order becomes curriculum-embedded, standardized and continua/MCT are next and roughly equal, followed by special school tests.

As was the case with the elementary teachers, while unavailability of certain kinds of assessment information early in the school year is perhaps to be expected, it is more surprising that so many teachers report non-availability once the school year is underway and decisions about instructional and classroom management modifications are being made. In this regard, about 10 percent of the math teachers report that no standardized test data are available; roughly 15 percent of the English teachers and 30 percent of the math teachers report that information from

district continua or minimum competency tests is not available to them; almost 15 percent of the English teachers and 25 percent of the math teachers report non-availability of information from curriculum tests; one quarter of the English teachers and almost 50 percent of the math teachers report the same for special tests developed or chosen by the school.

With regard to making decisions about students' report card grades, results of their own tests and direct observations of students remain of greatest importance for most secondary teachers. Results of curriculum tests appear next in order of importance as reflected by percentages of teachers, followed by results of tests developed or chosen by their school.

As was the case with the elementary teachers, note that the indices of non-availability of information for a given measure remain fairly constant between decisions involving student changes and decisions about their report card grades. That is, where information is reported unavailable for teacher decisions during the school year or semester, it also appears to be equally unavailable at or near the end of the year/semester. Perhaps for some teachers these measures simply do not exist; for others it may be (as seen later) that the results of certain measures which teachers have administered are not made available to teachers when they are needed for a given decision; perhaps the results of some tests are filed centrally and are never provided to teachers.

Summary

While we have seen that teachers' self-made tests and classroom observation are of great importance to teachers, many other kinds of assessment information are also important in their decision making.

Alternate sources of information, when examined as complementary tools set against decisions which are linked in the logic of the classroom, are seen in a less adversarial light than other work has implied. Teachers refer to many sources of information--perhaps too many but through no fault of their own--as they make the decisions they have to make; they want these decisions to be as informed as possible. Equally important here is that many of the information sources important to some teachers are simply not available to all teachers.

Although previous work has suggested some of the reasons why teachers use or do not use information, there may be other, perhaps more compelling reasons. Kinds of decisions to be made and the kinds of assessment tools made available, for example, may influence the relative values teachers place on information. Under the general theme of school or district characteristics bearing on test use, I'll try to develop other reasons from our data.

SCHOOL/DISTRICT CHARACTERISTICS BEARING ON TEST USE

One block of items on the national survey asked teachers about kinds of resources typically available to them. Questions in this series dealt both with instructional options and with options concerning tests or testing. Table 3 following presents the results and are listed separately for elementary and secondary teachers and for reading/English and math. The data represent percentages of teachers stating the frequency with which resources are available and used.

Instructional Resources

Although I do not emphasize purely "instructional" resources in this paper, let me make a point or two about these resources before discussing the "tests/testing" resource availability.

First, with the exception of instructional machines, in every other instance many more secondary teachers report non-availability of the resource than do elementary teachers.

Second, the resource option of alternative materials for independent work is the only resource which is available to almost every teacher.

Third, there is a marked difference between the number of elementary teachers for whom an outside specialist is available and the number of secondary teachers for whom this resource is available; this is especially the case with secondary English teachers.

Fourth, whereas 40 percent of the elementary teachers have some help from another adult or can work to some extent with another teacher, these options are not available to the vast majority of secondary teachers.

Finally, the data seem to paint a picture which, with the exception of the availabiity of elementary reading or math specialists, suggests that instructional resources, when they are available, consist largely of machines and printed materials and less of human resources. In most respects, the picture of test/testing resources, which is the interest of this section, is equally bleak.

Tests/Testing Resources

The data in Table 3 suggest that in terms of resource availability for tests and matters relating to testing, the option of working with other

Table 3

Resource Availability

Percentages Stating the Resource is Available

	Reading/English								Math								
	not available 0	available but not used 1	perhaps once a year 2	several times a year 3	about once a month 4	about twice a month 5	about once a week 6	more than once a week 7	not available 0	available but not used 1	perhaps once a year 2	several times a year 3	about once a month 4	about twice a month 5	about once a week 6	more than once a week 7	
Instructional Resources:																	
Another adult under my supervision (aide, volunteer, etc.) for small group or individual work	58	10	2	2	1	1	6	19	59	11	2	2	1	2	8	16	Elementary
	82	6	5	1	1	1	1	2	81	8	3	3	-	-	-	4	Secondary
One or more teachers with whom I divide students for extra help	48	9	4	4	1	1	2	31	55	11	5	3	1	1	4	21	Elementary
	76	10	4	4	-	-	2	4	74	13	4	4	1	-	1	3	Secondary
Instructional Machines (audio visual, computer terminals) for independent work	33	8	5	14	5	6	14	14	38	11	5	13	4	6	12	10	Elementary
	36	7	14	23	7	8	5	1	45	15	10	16	2	1	5	5	Secondary
Alternative published or teacher-made curriculum materials for independent work to meet special needs	2	1	-	11	3	6	20	56	3	1	1	11	6	6	20	51	Elementary
	12	4	6	21	10	6	18	22	14	10	6	17	10	10	17	16	Secondary
Specialists outside my classroom to whom I can send my students for special work	25	7	4	4	1	1	5	54	43	7	3	1	-	1	5	40	Elementary
	55	6	11	14	4	1	5	4	58	14	5	13	1	1	3	4	Secondary
Test/Testing Resources:																	
Someone who helps read, correct, or grade the tests and other assignments I give to evaluate students	69	5	2	4	1	2	6	10	69	6	1	2	1	1	7	13	Elementary
	70	4	-	3	3	3	8	9	70	6	1	3	-	2	6	12	Secondary
Other teachers with whom I plan and develop tests or other evaluation assignments	37	12	11	15	3	5	7	9	38	13	10	17	4	4	6	8	Elementary
	19	11	16	29	7	9	3	5	23	9	11	29	8	1	10	8	Secondary
Quick, computerized scoring and analysis of tests	64	2	21	8	2	1	-	1	64	3	21	9	1	1	1	-	Elementary
	58	16	5	15	3	1	1	-	59	12	8	17	2	1	-	-	Secondary
"Item banks" of test questions upon which I draw in making up my tests	72	4	1	6	5	4	5	2	70	3	2	9	4	4	6	2	Elementary
	51	8	9	15	8	3	3	2	53	9	3	13	4	6	7	4	Secondary

teachers in a test planning or development effort is the only tactic available to most teachers. Even then, only about 10 to 15 percent of the teachers report doing this fairly often.

In the other three areas--someone who helps read or score tests and other assignments; quick, computerized scoring and analysis of tests; and "item banks"-- most teachers simply do not have this resource. In this regard, the most extreme difference between the elementary and the secondary teachers is in "item bank" availability; many more secondary than elementary teachers have this option.

Note, once again, regardless of the resource being examined, that for those teachers for whom it is available only 10 to 15 percent report using it with any great degree of regularity. It seems likely, then, that although quite a lot of testing is going on, with a great deal of teacher reliance upon multiple sources of assessment information in their classrooms, the typical elementary or secondary teacher is virtually unassisted in terms of formal resource support.

Let me now take up the matter of the kinds of assistance provided to teachers by the school or district to help them make sense of the testing activities they are involved in.

District or School Assistance with Testing Activities

Tables 4 and 5 present the elementary and secondary teachers' responses to survey items dealing with this matter. In both tables, data represent percentages of teachers responding; in Table 5 separate data are shown for English and math teachers.

Table 4

District/School Assistance in Testing: Elementary Teachers

Responses reported in percentages

	Teachers Receiving This Assistance		Relevance for Classroom Work	
	NO	YES	Very Relevant or relevant	Slightly relevant or not relevant
How to administer tests required by my state, district, and/or school (procedures to follow, etc.)	22	78	67	11
Analysis and explanation of state, district, or school test results	16	84	72	12
How to construct or select good tests	80	20	17	3
Alternative ways (other than tests) to assess student achievement	46	54	45	9
Presentation of published materials designed to prepare students for particular tests or to improve test-taking skills	59	41	36	5
How to interpret and use results of different types of tests (e.g., norm-referenced and criterion-referenced tests and their applications)	41	59	49	10
How to tie what is taught more closely to the skills, content covered on required tests	50	50	42	8
Training in the use of test results to improve instruction.	65	35	29	6

Table 5

District/School Assistance in Testing: Secondary Teachers

Responses reported in percentages

	Teachers Receiving This Assistance		Relevance for Classroom Work		
	NO	YES	Very Relevant or relevant	Slightly relevant or not relevant	
How to administer tests required by my state, district, and/or school (procedures to follow, etc.)	46 54	54 46	38 30	15 15	English Math
Analysis and explanation of state, district, or school test results	30 40	70 60	55 43	14 17	English Math
How to construct or select good tests	77 82	23 18	18 11	5 6	English Math
Alternative ways (other than tests) to assess student achievement	75 79	25 21	19 16	5 5	English Math
Presentation of published materials designed to prepare students for particular tests or to improve test-taking skills	68 71	32 29	23 20	8 9	English Math
How to interpret and use results of different types of tests (e.g., norm-referenced and criterion-referenced tests and their applications)	65 66	35 34	28 22	6 12	English Math
How to tie what is taught more closely to the skills, content covered on required tests	63 75	37 25	30 22	6 3	English Math
Training in the use of test results to improve instruction.	79 81	21 19	16 14	4 5	English Math

For the elementary teachers, note that most of the respondents receive assistance in administering required tests, and in the analysis of state, district, or school test results. From that point on, the assistance drops of markedly. To be sure, more than half the elementary teachers report that they receive some assistance in the interpretation and use of different kinds of tests, and in alternative ways to assess student achievement. However, the vast majority report no assistance in the construction or selection of tests; this finding has a bearing on the possibility of teacher-driven criterion-referenced test construction and use. In addition, the assistance that is provided, limited as it may be, does not seem to emphasize the classroom uses of tests.

I mentioned earlier that a useful vantage point from which to view teachers and testing would be in assessment-instructional linkages. Two of the items on our survey tapped this potential--the last two items on Tables 4 and 5. Note that half the elementary teachers receive some kind of assistance in tying their teaching to required tests, but that two-thirds of them receive no assistance in using test results--of whatever form--to improve their instructional programs.

As a final point here, note that of those teachers who do receive specific assistance, most find it relevant to their classroom work.

Depressing as the picture may be for elementary teachers, it is even more so for the secondary teachers. First, once again it is only in matters relating to required or externally sanctioned tests that sizeable numbers of secondary teachers receive school or district assistance. In terms of test construction or selection, alternative assessment

possibilities, and interpretation and use of various assessment techniques, most of the secondary teachers, be it in English or math, receive no formal assistance. Further, as was the case with the elementary teachers, ways to foster assessment-instructional linkages are not provided to seecondary teachers.

In both samples, where assistance is generally provided, it is in the matter of required, externally sanctioned tests or testing programs. I will try to provide some reasons for this phenomenon later in the paper. In the next section I'll address the matter of district or school uses of assessment information.

District Uses of Assesment Information

Table 6 presents teachers' responses to a series of survey items asking how the school uses assessment results. These kinds of uses, on the one hand, get at whether the administration attempts to use assessment data to provide links with instruction, as in review of scores to identify instruction areas needing emphasis. On the other hand, they get at whether the administration uses test data in ways which might suggest to teachers that the data are being taken seriously, as in following up to ascertain whether teachers do emphasize needs identified by test scores, or requiring teachers to turn in scores on the tests they routinely give, or evaluating teaching or setting goals on the basis of test scores. I will amplify this matter of testing policy in a later section.

Clearly, these administrative uses of assessment data do not happen routinely for most teachers, whether secondary or elementary. Indeed, for two of the uses that might suggest the district's posture on the importance

Table 6

Teachers' Reports on the Extent to Which the School Makes Use of Test Results

Percentage of teachers reporting the activity

USE OF RESULTS	Does Not Happen at All		Happens rarely		Quite frequently but not regularly		Happens routinely	
My principal (or the school administration) ...	Elementary	10th Grade	Elementary	10th Grade	Elementary	10th Grade	Elementary	10th Grade
...reviews test scores to identify skill or content areas that need extra emphasis	13	36	18	37	21	14	38	13
...checks that I am emphasizing the areas identified by test scores as needing it	20	28	25	34	23	16	32	22
...requires me to turn in the scores or grades on the tests that I routinely give my classroom	64	71	12	12	6	6	18	11
...evaluates my teaching on the basis of test scores and/or establishes specific test-score goals for my students and me to meet.	72	84	15	6	8	8	5	2

of assessment data--turning in test scores and using test scores as part of teacher evaluation and/or goal setting--they do not happen at all for most teachers.

One final group of items reflecting relevant school or district characteristics remains to be discussed.

District Reporting of Test Results

The group of survey items in this series was concerned with test turn-around time and usefulness of test reporting formats. Of the elementary teachers, 46 percent indicated that test results are returned quickly enough so that they can potentially be used for instructional modification. Another 40 percent responded that results are received too late for this purpose. The remainder do not receive the scores back from the district. For those elementary teachers receiving the test results, while most found the format facilitates their use, about one quarter of them did not.

Of the secondary teachers, only about one quarter of the English teachers and of the math teachers responded that test results are returned quickly enough to be of use. About 35 percent of the English teachers and 25 percent of the math teachers indicated that the results are returned too late to be of use to modify instruction. About 10 percent of each sample responded that scores are not returned to them, and the remainder that the question does not apply. Of those secondary teachers receiving test results, opinion is just about equally divided as to the appropriateness of their format.

For a sizeable number of teachers, then, results are returned too late to be used in modifying instruction. In addition, and this is especially true for secondary teachers, their format has doubtful relevance.

INTERIM SUMMARY

There are many implications to be drawn from the findings of this paper and recommendations to be cast for practice and policy. Before I move into these matters, I'll briefly summarize what might be said about the current test-use picture. Then I'll describe an alternative for future testing practice based on our test use studies and the work of a few pioneering districts around the country. This work has a bearing on the remainder of the paper.

The Current Picture

Some of the previous work in test use and the secondary examination of that work has allowed the focus to drift. Though ostensibly describing uses of standardized, norm-referenced tests, the criticisms of these tests are frequently not legitimate because the tests in question are discussed and reported on in such a way that they appear to have a seemingly infinite range of legitimate functions. Many of these perceived functions of a norm-referenced test are contradictory and, hence, create a host of weaknesses competing for ascendency depending upon the particular test function under discussion. Further, given a wandering focus, the criticisms of standardized tests frequently sound like criticism of formal testing in general.

At the same time, some of the studies seem to suggest teachers need to make or that teachers say they make decisions which are discrete and linear and that they do this or would like to be able to do this on the basis of one source of information such as a norm-referenced test. Perhaps a more accurate picture of classroom practice would suggest that teachers are constantly making instructional decisions, many of those decisions overlap in purpose and in time and are cumulative, and hence teachers rely on a range of (sometimes overlapping) kinds of information. In this view, no single measure does, or should, emerge as the dominant, sole source of information. In this view, teachers' perceptions of the values of different kinds of tests are more evenly distributed, with one or two measures assuming fairly constant importance and, of the remainder at teachers' disposal, their weights of importance vary to the extent they can serve a teacher's decision-making purpose. I have tried to show that the tests and related resources available are not evenly distributed among teachers.

Some of the past studies of testing suggest that most criticisms of (standardized) tests are quite technical, relating to validity, reliability, breadth, narrowness, and there is no doubt that many of these criticisms may be fair from the perspective of an individual test user with a particular set of informatin needs; after all, standardized tests are intended to serve general rather than particular assessment needs, a point seldom given sufficient attention. But other criticisms, equally compelling, may apply. For example, in addition to unequal distribution of kinds of tests, many teachers receive no formal assistance in the appropriate uses of different tests, how they are to be interpreted, and especially how they can be of use in instructional planning and modifica-

tion. Many teachers receive the results of the tests they give too late and in a format inappropriate for use in their classrooms. In addition, most of the testing assistance they do receive is limited to tests the district requires, or is itself externally required to administer. Finally, most teachers may see their districts' failure to pay attention to classroom uses of information as an implicit measure of the role of testing in district policy.

In short, while I agree with earlier work that teachers have already offered many legitimate criticisms of tests and testing, other criticisms may stem from school or district uses of tests, assistance provided in testing and related matters, and coherence of testing policy, as perceived by teachers.

In addition, while I agree to some extent with what previous work has said about some of the uses teachers make of tests, I believe our work suggests that teachers use or at least refer to tests more than we had suspected and for a greater range of decision purposes. I mentioned earlier that teachers, through no fault of their own, are perhaps referring to too many kinds of test information. I suggest this happens because teachers are seldom provided with any assistance on the focal relevance of tests and testing, and that the most relevant focal point for teachers lies in the uses of assessment information for classroom practice.

Alternative for Future Practice

In test use and other work at CSE we have begun to identify some districts whose policy toward and uses of tests and testing differs markedly from the dominant mode and recognizes the relevance of testing for classroom practices. These districts are making serious attempts to link testing and evaluation information with instruction. In an earlier paper

(Burry, et al., 1981b) I described one of these districts as part of the exploratory fieldwork preceding our test use national survey.

I described this district as being loosely coupled in some regards and more tightly coupled in others. I'll amplify this distinction later. This variable postures appears to lend itself to multiple and complementary uses of assessment information: uses which are centralized and concerned with external accountability and reporting requirements and uses which are spread out and reflect the decision needs of individual schools and classrooms. This approach evolved over time in this particular district, and it seems to reflect not only the organizational reality of schools and districts but the careful determination of various decision needs and specification of an assessment information system that will meet these needs.

Assessment programs often intend to provide information for use at local, state, and/or federal policy levels. This can cause the program to emphasize, or to be seen as emphasizing, the information needs of one of these levels to the exclusion of others. As suggested in the findings discussed here, teachers might believe that the overall testing program is emphasizing external audiences and largely ignoring instructional uses of test data. Audiences associated with external requirements often ask for general assessment information that can be used to compare educational programs rather than more specific information to show the growth of individual pupils on a specific set of educational objectives. School systems responding more to the external audience than to others usually rely on the collection and analysis of pupils' scores on norm-referenced

tests. Teachers may get the impression that their schools are not overly concerned with assessing individual students and their growth in a given classroom. School systems responding more to their own internal audiences (and few seem to exist) might tend to rely more on criterion-referenced or objectives-based tests or teacher observation to provide information for diagnostic and prescriptive purposes. But a school system taking this position might be subject to questions about the educational significance of the scores obtained on these locally relevant tests--What do they mean? Do they show whether the learning that has taken place is important or trivial? How do the scores obtained on these tests compare with the scores obtained on other kinds of tests?

A school system might attempt to reconcile both kinds of information needs, to examine its total assessment requirements and needs, to determine which kinds of information will address the range of needs, to decide which kind of measure is most appropriate for generating the information addressing a particular decision area, to specify for its participants the intended uses of various measures, and thus design a coherent assessment program which is perceived to have a variety of overlapping uses.

One of the districts we did fieldwork in appears to have developed this kind of assessment program. It does this by establishing broad policy for the schools, and the schools in turn set policy for the instructional teams in the elementary schools and the departments in the high schools. In addition, both the district central office and the schools provide active leadership in the development and selection of tests and their instructional uses. Policy is clear, though flexible; and a great deal of

the testing appears to be "owned" by the school unit of concern--team or department.

Teacher knowledge of tests and testing has come to be quite sophisticated in the district through inservice and technical assistance that is largely provided by local school and district personnel. The testing situation appears to come close to the ideal. That is, it

- is parsimonious
- offers tests oriented to classroom teachers
- shows teachers how to use tests so as to meet their classroom instructional needs
- does not force teachers to emphasize tests that do not fit their practical demands
- permits teachers to administer/use a variety of tests
- is sensitive to the practical matters of teaching
- does not over-emphasize external reporting requirements, yet meets those requirements

In this district, the teachers, principals, and district officials seem to accept the need for and value in generating information that will paint the big (norm-referenced) picture, that will provide a wide angle view about groups and programs. They don't over-emphasize this picture. They also accept the need to generate information (criterion-referenced, or objectives based, or teacher observation) about individual students and classrooms that make up the big picture. They don't over-emphasize the value of this picture either.

They seem to be using the right kind of test to get the larger aggregate picture, and a series of other equally appropriate measures, with a different focus and with greater detail, to get a variety of snapshots that are more finely grained than the broader composition. The district has supplied the camera to get various pictures and takes the kind of shot with the degree of resolution it needs. The schools and classrooms use the same camera, but they select a kind of film that meets their needs, and then choose a speed, angle, focus, and degree of resolution sensitive enough to get the series of shots that they need. The end result is a montage reflecting different aggregates of students accomplishing a variety of tasks over time.

Other CSE work describing school and district attempts to link assessment and instruction is described by Bank and Williams (1981).

IMPLICATIONS AND RECOMMENDATIONS

In the remainder of this paper I'll try to draw some implications, both from CSE test use and other data, for schools who may wish to establish an instructional focus for at least part of their assessment programs. Where I can, I'll offer some tentative recommendations.

There appear to be at least three kinds of potential barriers--methodological, technical, and organizational--in the way of establishing assessment-instructional linkages.

Methodological Considerations

In a recent CSE monograph, Resnick (1980) describes the domineering

influence of developmental and differential psychology on education. Neither school of thought believes very strongly in the power of education--specifically instruction--to influence children's capabilities. Resnick describes developmental psychology as offering a theory of natural development which is more efficient in suggesting how not to interfere with development than in how to promote development. She pictures differential psychology as useful in describing and classifying children and as a discipline which sees education as adapting to children's capabilities rather than creating capabilities. These dominant forces have strongly influenced our disposition toward instruction and its assessment; their impact is seen in the findings previously discussed.

Resnick suggests that as education attempts to develop competence for all children, our traditional reliance on the dominant psychological models will need to be lessened in favor of an increased reliance on learning psychology. She describes learning psychology as believing more in the power and potential of instruction and, indeed, as embodying more knowledge about how to design instruction.

If we are to address the issue of education's technical core--instruction and its value--then we first need to examine more closely, and perhaps alter our views about, the psychological models that influence our views of schooling.

As part of our work examining the role of school district evaluation offices--where they are, how they are staffed, what they do--we commissioned a series of papers to re-analyze the empirical data we collected, test the data against some theoretical propositions about

evaluation and assessment in educational organizations, and cast some recommendations about the need for changes in evaluation and assessment.

In one of the papers, O'Shea (1981) provides a politico-historical explanation of why district evaluation units are more likely to engage in achievement monitoring--did the program accomplish what it intended? as opposed to analytic evaluation--what was the worth or value of the program's accomplishment? This work is philosophically related to the view that learning psychology has a powerful role in the design of instruction, and it suggests that with the use of alternative forms of evaluation and assessment we can promote linkages between assessment and instruction.

O'Shea's concern was to describe those factors inhibiting analytic evaluations in schools and districts and to suggest what might be done to facilitate analytic evaluations. His principal point is that evaluation (as defined above) of instructional programs as opposed to their monitoring is thwarted by contradictions between the assumptions guiding most evaluations and the nature of instruction in school settings.

He views the major inhibiting factor as stemming from the dominance of the experimental paradigm in most evaluations. Evaluation following this mode, with its assumptions about treatment cause and effect, obeys the logic of technical rationality, which schools do not fit as we presently have little theory from which to specify which instructional opportunities will lead to specific learning outcomes.

In addition, O'Shea continues, schools are institutional rather than technical organizations. In a technical organization, where means-ends

relationships are known, evaluation can determine the efficiency with which ends are achieved and the benefits of those ends in terms of a certain set of specified costs. (Ways of beginning to identify and weigh these costs are offered by Catterall later in this report.) Institutional organizations, on the other hand, do not have a well articulated technical core in which instructional cause and effects are known, and thus most assessment programs monitor achievement of stated outcomes, usually on the basis of some norm-referenced measure, rather than attach worth to them by measuring individual pupil growth toward maximum potential.

Schools and districts wishing to examine relationships between instruction, assessment, and learning will need to adopt evaluation methodologies and assessment devices not at loggerheads with the nature of schools as institutional organizations. These methodologies will need to be informed less by the experimental paradigm and more by the qualitative methods of ethnomethodology and anthropology, and enhanced by measures, such as criterion-referenced tests, which permit examination of individual pupil growth rather than provide some (exclusively) norm-referenced view of the program in the aggregate. These more naturalistic observations may begin to suggest the outlines of instructional cause and effect relationships.

These views of Resnick and O'Shea tend to emphasize the potential power of instruction, a power which teachers in our study would probably be interested in given their concern for instructionally related assessment information. Attempts to supply teachers with the appropriate tools, however, will need to be part of a larger effort which addresses the over-arching methodological considerations.

I am not suggesting that either the experimental mode of investigation or the norm-referenced approach to assessment be abandoned; only that they be placed in a context which recognizes the legitimacy of other approaches, depending upon the specific evaluation and assessment tasks at hand.

Technical Considerations

The data in our test use study offer teacher-perceived technical limitations of tests, and also suggest that teachers might view their district testing policy as having little coherence, especially from the standpoint of how testing ties in with instruction and the importance the district appears to place on instructionally-linked assessment in relation to other district needs. Therefore, any attempt to work with teachers in the appropriate uses of tests will need to address both test property and test relevance. Teacher training in the former should deal with a test's psychometric properties, the assumptions which drove its planning and construction, its legitimate uses, and its instructional applications. CSE is addressing such problems as those associated (1) with describing a test's properties and assumptions, (2) training teachers and others in test development and selection, and (3) administering and using these tests for instructional purposes. We have already developed training materials for these purposes (Baker, Polin, & Burry, 1980). These materials get at what seems to be a central concern for teachers describing test use; that is, selecting or developing tests which meet teacher concern for validity, therefore, which involves test match not only with what is taught, but

also with how it is taught and to whom. This kind of training seems a logical point of entry for districts attempting assessment-instructional linkages, especially in light of the limited training provided to teachers in test development and test use.

To be sure, there are higher-order measurement considerations to be kept in mind, such as whether notions of classical test theory will transfer and apply in the realm of criterion-referenced testing. But that is not the focus of this paper since it need not become an issue in which teachers will be embroiled.

Test relevance training should address teachers' concerns with test purpose, focus, and use; for example, through the establishment of a testing policy that has coherence and usefulness for teachers. Planning such a policy and testing program should not be taken lightly since it involves not only providing people with knowledge of tests and testing but must also deal with attitudes towards tests and testing. Planning a testing program that attempts to balance the need to generate information for external reporting requirements and information for internal instructional decisions will not be all that easy. However, the practices alluded to earlier that are already taking place in a few districts addressing assessment-instructional linkages will offer some initial starting points.

The matter of how teachers might continue to use their own measures, and how these measures might acquire a more secure footing in district policy, is still tricky. But CSE has begun work with teachers and district staff in how teachers' own assessment techniques can be used in such a way as to preserve their classroom instructional relevance at the same time as tying in to larger evaluation considerations.

Training people in test selection, test development, the use of qualitative measures, and instructional uses of assessment will be a difficult task. To supplement CSE training materials in these matters we have begun to develop a series of resource papers for the practitioner (Burry, 1981a, 1981c; Baker, 1981; Herman, 1982).

The primary job, it seems to me, will be to make efforts to involve central office staff, school administrators, classroom teachers, and others, through training and other resources, in concerted and collegial planning of a testing program which addresses a variety of needs. The difficulty, however, may be in overcoming some organizational problems which might make such planning difficult.

Organizational Characteristics

Any attempt to establish a testing program and a surrounding policy that is mutually acceptable to central office staff on the one hand and to principals and teachers on the other must address the organizational realities of schools.

In one of the Kappan articles I mentioned earlier, Sproull and Zubrow (1981) find that testing is not an important consideration for most central office administrators. Our findings would suggest one or two important qualifications to that conclusion, and would suggest that administrators might not consider testing, except for external reporting needs, as important; or, from the standpoint of teachers, it might appear that administrators do not consider testing, except for external reporting requirements, as important. To the extent that either statement applies in

a school district, then the need for collegial planning and establishment of a coherent assessment program becomes critical.

Research and development specialists, evaluators, teachers, test developers, curriculum specialists, and administrators need to work together to formulate ways in which evaluation and assessment information can be used to meet a variety of needs which complement, rather than confound each other. A major concern of this effort will be to bring to bear on the problem a variety of specialties and points of view to ensure that testing and instructional matters are considered in concert, that external and internal assessment and reporting requirements are balanced. In this regard, just as teachers cannot be faulted for the range of testing expectations they express, neither can administrators really be found wanting, given their organizational realities, if they stress, or are perceived to stress, the uses of assessment information for external audiences.

I already alluded to some of the work we commissioned (O'Shea, 1981) which has a bearing on this issue. Some of the key terms already offered were institutional vs technical organization and loose or tight coupling. Other papers in the series (O'Reilly, 1981; Zucker, 1981: Grusky, 1981) elaborate these issues.

Schools and districts face several organizational dilemmas. First, they are institutional organizations. As such, they are held accountable to society, often via fu ·ing agencies, to meet societal expectations and their evaluators and testing specialists may feel their principal mission

is to justify what the schools do for audiences external to the school system. Further, unlike technical organizations--automobile producers, food wholesalers and retailers, appliance manufacturers, whose output can be measured directly against given input--institutional organizations lack a strong technical core. In our case, technical core means instructional treatments and specified outcomes. With these two organizational features in mind, educational evaluation, and the assessment practices it relies upon, is (1) primarily directed to the needs of external audiences, and (2) even if it began to focus on more internal--technical or instructional--decision needs, the necessary theory of means-end is lacking. Without this theory, the technical core of education, instruction, is loosely coupled or decoupled from its surrounding organization, and evaluation practices and evaluation information are seen to have little relevance for those who provide instruction.

Because of the above considerations, evaluation and testing people do not enjoy a great deal of status within the educational organization--especially with regard to instruction. Should schools make systematic attempts to link assessment and instruction, they may find themselves competing for resources and recognition with central adminstrators and evaluators concerned with external reporting needs. Once again, the system-wide planning of an overall evaluation-assessment program, with multiple purposes, will be critical. Part of this planning should consider how evaluation and assessment functions can focus inwardly as well as outwardly, and how the two can be legitimately placed and recognized in district policy.

I think that we can take hope from the few districs attempting assessment-instructional linkages. CSE test use data from teachers suggest their need for instructionally related information. Earlier in this report Choppin and Dorr-Bremme offered suggestions that might be used not only to reduce the amount of testing taking place, but also to increase its usefulness to a variety of audiences with different information needs. With the proper approach, teacher need might be linked with change in assessment practice and policy which would not only enlarge the constituency to be served but might also shed some light on the questions we have about education's technical or instructional core.

REFERENCES

Airasian, P.W. The effects of standardized testing and test information on teacher's perceptions and practices. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979.

Airasian, P.W., Kelleghan, T., Madaus, G.F., & Pedulla, J.J. Proportion and direction of teacher rating changes of pupils' progress attributable to standardized test information. Journal of Educational Psychology, 1977, 69(6), 702-709.

Baker, E.L., Polin, L., & Burry, J. Making, choosing, and using tests; A practicum on domain-referenced testing. Center for the Study of Evaluation, University of California, Los Angeles, 1980.

Baker, E.L. Achievement tests: What they are and what they are for. CSE Resource Paper No. 3. Center for the Study of Evaluation, University of California, Los Angeles, 1981.

Bank, A., Williams, R., & Burry, J. Evaluation in school districts: An organizational perspective. CSE Monograph No. 10. Center for the Study of Evaluation, University of California, Los Angeles, 1981.

Bank, A., Williams, R., Evaluation Design: Conference Proceedings. Evaluation Comment, 1981, 6(3).

Bauerfeind, R.H. Building a school testing program. Boston: Houghton Mifflin, 1963.

Boyd, J., McKenna, B.H., Stake, R.H., & Yashinski, J. A study of testing practices in the Royal Oak (Michigan) public schools. Royal Oak, Michigan: Royal Oak City School District, 1975. (ERIC Document Reproduction Service No. ED 117 161.)

Broekhoff, R.H. Criterion-referenced evaluation as an alternative. English Journal, 1978, 67(7), 32-37.

Burry, J. An introduction to assessment and design in bilingual program evaluation. Bilingual Education Paper Series, 1982, 5(6). Also appears as CSE Resource Paper No. 1, 1981a.

Burry, J., Dorr-Bremme, D., Herman, J., Lazar-Morrison, C., Lehman, J., & Yeh, J. Teaching and testing: Allies or adversaries. CSE Report No. 165. Center for the Study of Evaluation, University of California, Los Angeles, 1981b.

Burry, J. Evaluation and documentation: Making them work together. Bilingual Education Paper Series, 1982 (in press). Also appears as CSE Resource Paper No. 2, 1981c.

Ebel, R.L. Improving the competence of teachers in educational measurement. In J. Flynn and H. Garber (Eds.), Assessing behavior: Readings in educational and psychological measurement. Reading, Mass: Addison-Wesley, 1967, 171-182.

Gorrow, F.F. Better classroom testing. San Francisco: Chandler Publishing Co., 1966.

Goslin, D.A. Teachers and testing. New York: Russell Sage Foundation, 1967.

Grusky, O. Role conflict and ambiguity among school district evaluation heads. In A. Bank, R. Williams, & J. Burry (Eds.). Evaluation in school districts: An organizational perspective. CSE Monograph No. 10. Center for the Study of Evaluation, University of California, Los Angeles, 1981.

Hastings, J.T., Runkel, P.J., & Damrin, E.E. Effects on use of tests by teachers trained in a summer institute. Urbana: University of Illinois, Bureau of Educational Research, 1961.

Herman, J.L. Criteria for reviewing district competency tests. CSE Resource Paper No. 4. Center for the Study of Evaluation, University of California, Los Angeles, 1982.

Howe, H., II. Test and schooling. Washington, D.C.: Paper presented at the National Conference on Achievement Testing, Washington, D.C., 1978. (ERIC Document Reproduction Service No. Ed 152 852.).

Klein, S.P. Evaluating tests in terms of the information they provide. Evaluation Comment, 1970, 2(2), 1-6.

O'Reilly, C. Evaluation information and decision making in schools: Some constraints on the utilization of evaluation information. In A. Bank, R. Williams, & J. Burry (Eds.). Evaluation in school districts: An organizational perspective. CSE Monograph No. 10. Center for the Study of Evaluation, University of California, Los Angeles, 1981.

O'Shea, D.W. School district evaluation units: Problems and possibilities. In A. Bank, R. Williams, & J. Burry (Eds.). Evaluation in school districts: An organizational perspective. CSE Monograph No. 10. Center for the Study of Evaluation, University of California, Los Angeles, 1981.

Perrone, V. Remarks to the National Conference on Achievement Testing and Basic Skills, Washington, D.C., March 1978.

Resnick, L.B. Social assumptions as a context for science: Some reflections on psychology and education. In H.D. Gideonse, R. Koff, & J.J. Schwab (Eds.). Value, inquiry, & education. CSE Monograph No. 9. Center for the Study of Evaluation, University of California, Los Angeles, 1980.

Resnick, L.B. Introduction: Research to inform a debate. Kappan, 1981, 62(9), 623-624.

Salmon-Cox, L. Teachers and standardized testing: What's really happening? Kappan, 1981, 62(9), 631-634.

Sproull, L. & Zubrow, D. Standardized testing from the administrative perspective. Kappan, 1981, 62(9), 628-631.

Stetz, F. & Beck, M. Teachers' opinions of standardized test use and usefulness. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, 1979.

Zucker, L.G. Institutional structure and organizational processes: The role of evaluation in schools. In A. Bank, R. Williams, & J. Burry (Eds.). Evaluation in school districts: An organizational perspective. CSE Monograph No. 10. Center for the Study of Evaluation, University of California, Los Angeles, 1981.

THE COSTS OF SCHOOL TESTING PROGRAMS

James Catterall

INTRODUCTION

Like the horse and carriage, schooling and testing go together. From the surprise quiz to the competency exam, assessments in the form of tests are universally practiced in the schools--as integral parts of curricula, as guides to pupil placement, and as indicators of educational health. But despite recent obervations that testing is proliferating in American schools (Reznick, 1981) and that statewide testing programs often command sizeable budgets (Anderson, 1977), neither education researchers nor policy analysts have yet taken a comprehensive look at the costs of testing in the schools, or even described how such a task might proceed.

This paper represents a preliminary inquiry into the costs of testing in elementary and secondary schools. Our primary purpose is to create ways of thinking about the topic, since there are no cost paradigms that have established a permanent home in the vast literature on testing. Our efforts will serve more to provide an underlying framework for substantive research about costs and testing, than to provide immediate empirical conclusions about the magnitude of current testing costs. Given the present state of the art, our cautious construction of these foundations seems warranted.

We begin with a theoretical model which captures certain critical relationships concerning the costs and benefits of testing. Introduced as an economics of information paradigm, the model regards our interest in the topic of costs and testing to include construct of optimality in the amount of testing conducted in schools and efficient use of testing resources. Both of these ideas demand precise knowledge about the costs of testing. We then discuss the fundamental elements of any analysis of testing costs--namely, identifying and evaluating the costs of tests or programs under scrutiny. These first steps will be seen to apply to cost analyses performed under our economics of information paradigm, and to cost analyses performed according to more familiar analytical frameworks such as cost-benefit analysis and cost-effectiveness analysis. The heart of our discussion remains with the issues surrounding, locating, and estimating testing costs because of the importance of these tasks to all "higher" forms of cost analysis. We conclude with a discussion of the implications of our remarks for substantive research into actual testing costs.

The Economics of Information and Testing

School professionals and education researchers hardly need to be reminded that information is a valuable resource. In part our schools exist for the purpose of transmitting knowledge--i.e., information that is implicitly held to be of value. And researchers (or their sponsors) pay dear prices for the information they collect in the name of educational inquiry. That information, like any good, has both value and cost and has led economists to the formulation of an economics of information paradigm (Stigler, 1961). The paradigm is not so much a sub-discipline within the field of economics as it is a way of applying neo-classical economic models and micro-economic reasoning to the phenomenon of information-seeking. The

paradigm addresses such questions as what amount of resources should a decision maker, such as a testing authority or a teacher, allocate to a search for information? Or put another way, what are the patterns of costs and benefits associated with information collection?

While the economics of information literature primarily addresses consumer behavior and market information (e.g., how long does one search for a lower price?), the overall perspective has direct applications to the phenomenon of testing in the schools. By its very nature a test is a device for collecting information. The information created by such assessments can be regarded to have value to any or all of a number of audiences--pupils, parents, teachers, administrators, public officials, and society. Testing also has both direct monetary costs which appear in school and district budgets and indirect opportunity costs which are reflected in the use of resources that are not specifically budgeted for testing. Figure 1 presents a typological outline of the costs (and benefits) of testing, and implies certain definitions and relationships that will contribute throughout the balance of this discussion.

What Types of Costs are Associated with Testing?

It is helpful to have concrete notions of costs and benefits of testing in the schools in mind before considering the application of any of our analytical constructs including the economics of information paradigm. Figure 1 represents an attempt to identify the various types of costs (and benefits) which can be associated with testing, the first step in cost accounting. Overall, Figure 1 illustrates the complexity of the general topic of costs and testing; it also points to the relationships which are helpful in our analysis. A few explanatory notes are needed:

Figure 1

Costs and Benefits of Testing: Broad Typology

I. Costs Potentially Related to any Test

- Development
- Administration
- Analysis of Results
- Dissemination of Results
- Psychological Costs (e.g. stress, self-image)

Cost Elements
Professional
time(oppty)
service ($)
Clerical
time(oppty)
service ($)
Pupil Time (oppty)
Materials ($)

II. Costs Related to Outside* Mandates

- Legislation
- Monitoring and Enforcing
- Compliance
- Avoidance
- Cost of Consequences (e.g., remediation or legal costs)

Policy Costs

Debatable status as costs "of testing"

III. Benefits of Testing (All ultimately tied to system effectiveness)

A. Information Benefits

- Instructional management
- Pupil administration and guidance
- Curriculum decision making

Higher-level policy making

B. Other Benefits

- Incidental learning
- Pupil motivation
- Institutional motivation
- Demonstration of concern for school performance
- School-community-parent communications

*"Outside" refers to levels above the teacher/classroom, most often district or state mandates.

Notes to Figure 1

- Section I of Figure 1 lists types of costs which may be associated with any test. All tests involve administration and analysis of results in some form. Development costs are relatively large for tests like new statewide mandates, and more negligible for a weekly algebra quiz, and so on.

- Each type of cost listed in both Sections I and II can involve a variety of tasks which are not specified. For example, test development can involve identification of objectives to be assessed, item construction, and designing and validating the testing instrument. Or legislative costs may have many components which are not specifically shown. The categories listed are potentially to be considered as umbrellas for multiple activities.

- Each of the types of costs listed in Sections I and II can generally be expressed in terms of the cost elements shown at the right margin. These may be direct dollar costs for personnel engaged or materials purchased, or they may represent opportunity costs such as the time of personnel already hired or the time of pupils. With respect to individual time, it is necessary to consider time both before and after test administration in addition to test administration time. (Preparation time and lost class time for "cooling out" test takers are examples.)

- The costs related to outside mandates (Section II) include a number of categories that are not relevant to normal curricular testing. Mandated testing programs must be conceived and legislated (sometimes including experimental studies or other research and analysis); they also must be implemented and monitored. Further, they impose costs of compliance and avoidance. Such costs pertaining to these mandates can be seen as opportunity costs since they are resources which could be devoted to other purposes within the educational or public sectors.

- Outside mandates may create costs because of their consequences, such as remediation costs for pupils who do not pass competency tests, or legal costs if public officials are sued as a result of the prospect or outcomes of tests. Whether these costs should be considered to be costs of testing is problematical, and analysts might recognize the need to establish boundaries which delimit costs that are attributed to testing.

- The benefits of testing listed are generally tied to goals of effectiveness within the educational system. The informational benefits when conferred--accruing to pupils, teachers, administrators, and policy makers at all levels--can be assumed (or hoped) to have an ultimate positive impact on instruction. Teachers can plan their lessons according to the information gained in assessments of their pupils, district officials can assign pupils to classes and programs appropriately, and public officials can create or modify programs according to what is revealed by tests. (This is not to imply that all tests in fact confer such benefits, which is an empirical question.)

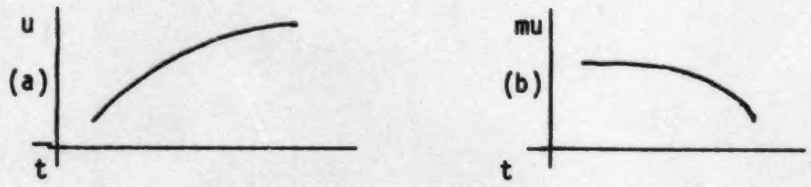
- There are also non-informational benefits which can be attributed to tests. Incidental learning through test taking is one probable benefit. Curricular tests can serve to inspire more diligent study (due to pupil desire for good grades, or to their aversion to failure), and state assessments can have the explicit purpose of improving the performance of schools. Also, political benefits may accrue to decision makers who adopt testing programs as a demonstration of their concern for education.

With this general inventory of costs and benefits of testing in mind, we will now consider their relationships according to the fundamental principles of the economics of information model. The discussion refers to Figure 2 which provides several illustrations. The economics of information paradigm suggests that certain basic relationships would hold between the amount of testing undertaken in the schools and both the costs and benefits associated with such testing. These relationships further imply that there exist, at least hypothetically, optimum levels of testing. In this analysis it may be useful to consider these relationships from the point of view of a single actor or office (for instance the classroom teacher), although they could also be extended to apply in other levels of analysis. Examples cited will adopt the narrower perspective.

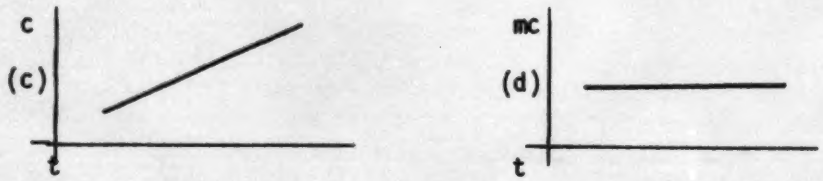
Figure 2

Relationships Among Amount of Testing, Costs, and Benefits

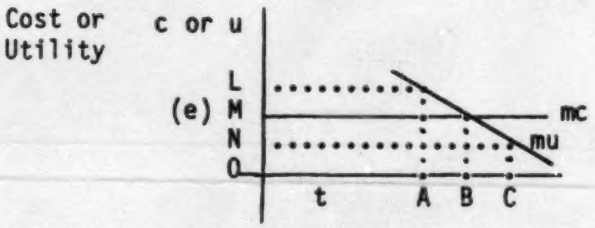
I) Diminishing Marginal Utility of Testing



II) Cost of Testing



III) Optimum Levels of Testing



t = amount of testing

The first principle shown is called the diminishing marginal utility of testing. This refers to the likelihood that beyond a certain point as a teacher gathers information, successive increments of information will be less and less valuable. As the teacher gains information about pupils in a regular testing program, we would expect added testing beyond some point to contribute less and less to the total usefulness of the information obtained. In section (I) of Figure 2, this is shown in two ways. In graph (a), u refers to the total utility (or usefulness or benefit) of information gained from tests, and (t) refers to the amount of testing conducted. While the total utility may continue to rise as more testing is done, the amount added to that total for each additional unit of testing steadily diminishes. Therefore, the curve becomes less steep as it moves toward the right. The adjacent figure shows that the added gain from testing or marginal utility (mu) diminishes as the amount of testing increases. Marginal utility is defined as the amount of utility added as a result of successive increments of testing. The shape of the marginal utility curve is derived from the shape of the total utility relationship to its left in the figure.

The second set of illustrations, II (c) and II (d), illustrate a hypothetical, but likely, cost relationship in testing. We assume that the costs of testing are approximately proportional to the amount conducted. This assumption is based on the fact that most of the types of costs listed in Figure 1--particularly those related to test administration and analysis--are directly tied to the amount of testing. This relationship is represented in graph (c). In the graphs, costs are shown to rise in direct proportion to the amount of testing. The marginal cost (mc) stays constant as shown in graph (d), since added units of testing are assumed to contribute to costs equally.

The importance and synthesis of these theoretical relationships pertaining to the costs and utility or benefits of testing are illustrated graphically in Figure 2 (e). A necessary assumption we make in synthesizing the costs and benefits for this illustration is that they both must be thought of in equivalent units of measure. The most relevant construct for testing in this regard is the notion that both costs and benefits might be expressed in terms of instructional effectiveness. Testing contributes to instruction in various ways and its costs ultimately (although sometimes remotely) represent other learning opportunities foregone. Resources taken from testing--i.e., dollars, personnel, materials, or others--could find a variety of alternate productive uses. The linking of costs to instructional effects is consistent with the typology of costs presented in Figure 1, even though the precise relationships between such resources and instructional effectiveness remain unspecified.

The synthesis shown in Figure 2 (e) is best described in conjunction with a classroom example. Consider the teacher's decisions regarding an appropriate amount of testing. On the one hand, testing brings gains in the form of information (and perhaps incidental learning). On the other, it exacts a variety of costs. According to the model, added testing is a winning proposition up to a certain point, and a less favorable proposition beyond that point. If the teacher is conducting an amount of testing corresponding to point A in the illustration, increasing this amount of testing would bring relatively more gains than costs. Gains are read in the diagram as OL, since the marginal utility (mu) curve shows this to be the amount of gain associated with increments of testing at this level; costs are shown as OM, since the marginal cost (mc) curve depicts this to be the additional cost of added units of testing at this level. This

relationship, which urges more testing, holds up until the point where the added benefits of testing just equal the added costs--the point indicated by B where both added utility and added costs equal OM. Beyond this point B, the instructional effectiveness of additional testing is shown to diminish, overall, because the addition to benefits caused by added testing is less than the addition to costs. Point C illustrates such a condition.

The economics of information model presented serves more to organize certain thoughts about costs and testing than to provide a ready blueprint for empirical assessment. Its first suggestion is that both the costs and benefits of testing might be thought of in equivalent terms, i.e., their ultimate impact on instructional effectiveness. Then, given likely patterns of overall costs and benefits associated with differing amounts of testing, the model suggests that optimal amounts of testing are at least theoretically identifiable. The first suggestion provides guidance as to how the costs of testing might be usefully conceived. The second of these suggestions provides a basic rationale for an inquiry into the costs of testing, since the level of costs identified take on importance in the context of normative judgments about the amount of testing occurring in the schools.

Common Cost Frameworks and Testing

The economics of information paradigm encompasses certain more common cost analysis schemes which could be included in our discussion, but which will not be discussed for a variety of reasons. Our first objective is to create an overarching framework within which the costs of testing can be approached (the economics of information paradigm). Our second is to point to substantive first steps that can be taken by analysts who have an interest in pursuing empirical investigations of the costs of testing in

the schools. This task is addressed in the balance of this paper. Familiar frameworks, such as cost-benefit analysis and cost effectiveness analysis, while having obvious connections to the paradigm set out above, lie somewhere between these two objectives--and so they must fall to the dictates of priority and space limitation in this report. These constructs treat the costs and benefits or effects of testing in ways that are useful to specific investigations, and with individual limitations which must be recognized. But either of these types of cost analysis and, more important, any analysis that is proposed under the broader paradigm, must begin with the critical issues of identifying and evaluating costs, to which we now turn.

The Building Blocks of Cost Analysis: Cost Accounting

The first steps of any cost analysis can be called cost accounting. We will first discuss the ideas generally, and then apply them to an examination of testing costs. Cost accounting is the dual task of identifying all costs pertaining to a program or policy and evaluating the magnitude of each type of cost. Cost analysis conducted under any of the frameworks we have discussed must begin with these accounting activities. We cannot compare programs on the basis of costs, nor can we relate program outcomes to costs without first knowing the types and levels of costs associated with specific programs.

IDENTIFICATION OF COSTS

In practice, the identification of costs has both direct and obvious aspects as well as potentially important dimensions which can escape detection. The direct dollar costs of programs are patently visible to the

analyst, since they represent resources which must be produced in order to initiate and maintain an activity. As examples, we might consider the school district which is planning for a minimum pupil competency testing program and is examing two alternative approaches--buying a minimum competency testing package at a set dollar cost per pupil or hiring a consultant to develop a competency test for the district. The perceived costs associated with each choice may be limited to the dollar cost of buying the packaged tests in the first case or the size of the consultant's fee in the second case. A simplistic cost comparison might incorporate these direct costs and nothing more.

But a number of costs in educational programs do not represent direct cash outlays to their sponsors and are, therefore, easily overlooked in cost estimates and cost comparisons. These costs can be buried in the use of resources which already appear in the sponsor's budget--such as the use of teacher or clerical time. They also appear as costs that are borne by entities other than the sponsor, such as other agencies or private interests. These less direct costs are best understood in the context of the full range of types of costs attached to educational programs and with an understanding of who, including the sponsor, would be responsible.

Figure 3 illustrates both the range of types of costs which must be considered in identifying and evaluating the costs of a program, and also the various entities which might have to bear the burden of those costs. Of course, the specific characteristics of a program being examined will determine just which types of costs are relevant, and just which sources will "pay" each and to what degree. Our simple school district competency testing example can now serve us further. Beyond the cash costs for purchasing a test package or a consulting fee for test development, the two

Figure 3

Illustrative Framework for Cost Accounting In Educational Programs

Entity Bearing Costs

	Cost to Sponsor	Other Agencies or Levels	Contributed Private Input	Imposed Private Costs	Total cost
Personnel					
Facilities					
Material & Equipment					
Other (Specify)					
Value of Client Time and Other Client Input					

(Adapted from Levin, 1975, p. 101) TOTAL $_____

alternative testing strategies may involve various more hidden costs according to this framework. If the testing package under consideration for purchase includes scoring and reporting services while the consultant's plan involves district clerical personnel or teachers for test scoring and score analysis, the consultant's plan contains a hidden cost that is borne by the district--clerical and teacher time. If the consultant is provided without fee by the state education agency, the consultant option is free from the point of view of the school district, but it actually entails a cost that is borne by the state--an outside agency. And yet another ramification for cost analysis surfaces in this example: If two tests

being considered by a district require significantly different amounts of time for their administration, they exact differing amounts of additional valuable resources--teacher and pupil time. In the terms of Figure 3, these costs would come under the value of client time.

The costs within programs which do not involve cash outlays, but which do involve the reallocation of a sponsor's resources to projects under consideration, can be called opportunity costs. When resources are engaged in one activity, they are by definition unavailable for other tasks. When clerical personnel are assigned to test-related activities, they necessarily will utilize time which could be devoted to other purposes. And while personnel allocation in this fashion does not involve direct cost implications--employees are on the payroll regardless of their assignments --the school district sacrifices the use of these resources for other purposes when they are assigned to a particular program. Opportunities are thus foregone, engendering the term "opportunity costs" (which in the economists' vocabulary refers specifically to the value of the best alternative use for a resource).

Several entries in the framework shown in Figure 3 were not relevant to the competency testing example. This is consistent with the notion that the identification of costs and of agencies or individuals responsible for them, is specific to individual policies or programs. The remaining categories in the framework are nearly self-explanatory, but a few comments are offered: "Facilities costs," as with "personnel costs," may involve direct elements such as buying or leasing space, as well as the assignment of existing facility space to a proposed project--hence an opportunity cost. The "other cost" category allows for identified costs which do not fit elsewhere in the scheme--travel is one example. On the incidence dimension (across the top of the figure) "contributed private inputs" include services such as time donated by volunteers. Volunteer services are best understood as opportunity costs since volunteer resources are generally scarce and have alternative uses. "Imposed private costs" refer to such costs as pupil transportation when this is required by a program and then provided by the clients at their own expense.

Some Practical Issues in Test Cost Identification

We have maintained that all cost analysis paradigms require specification of the various costs embodied in programs. This is an immediate issue for current research simply because the literature does not offer a taxonomy of the costs of testing across the full spectrum of testing activities in the schools. Recall that Figure 1 offers an inventory of types of costs associated with testing. This inventory provides only a starting point in the cost-identification process since a variety of

activities with cost implications may be identified within each category presented. Also, the presence of a variety of cost elements within each activity adds further complexity to the task of cost identification. For example, test development involves a range of tasks including both analysis of the domain of subjects and skills to be assessed, and also creation of an assessment instrument including the development and validation of test items. The costs which might be associated with these activities are multiple.

The identification process will be specific to the type of test or tests being examined. Cost types and elements may or may not pertain to a given inquiry, depending on the type of testing involved. For instance, we could inventory all testing being conducted in a school, or in an "average" school, and proceed to tabulate for all associated costs. Or we might select a specific test or type of test, such as an annual district assessment, or a year's worth of unit tests in reading, and proceed to identify the costs associated solely with those assessments. Our object of inquiry dictates the specific costs to be included for analysis. In short, the identification process inevitably returns to the nature of the question(s) we are asking in the first place.

EVALUATION OF COSTS

The above paragraphs outline the tasks of cost identification--determining what elements contribute to testing costs. Once relevant costs are identified, and the sources of responsibility assigned, the costs must be evaluated to complete the tasks of cost accounting. As in the problem of identifying costs, the evaluation of costs has both direct and indirect

qualities. In the case of resources for which there is a competitive market, such as personnel, materials, and facility space, direct cost estimates can be obtained from examination of existing budgets or through cursory market surveys. Costs of these services must be accurately assigned to budget periods under consideration. For example, equipment costs should be amortized over their expected useful life in order to estimate an annual cost, if that is desired. And in cases where resources are devoted jointly to more than one program, their costs to a single program must be assessed on a share-of-use basis. The indirect costs--facilities that have been paid-off, volunteer time, and client time, for example--can be estimated on the basis of opportunity costs. There are a number of standard references to cost estimation and cost allocation which offer more detailed prescription in these methods of cost assessment and allocation than we will provide here (Horngren, 1967; Anthony, 1964).

Evaluation of Testing Costs--Practical Issues

The evaluation of the costs identified in a particular inquiry has been presented as a critical second step in any cost analysis. After enumerating the costs, questions of magnitude arise. The inventory presented in Figure 1 reveals several types of costs which might have to be evaluated. Some of these costs can be immediately linked to dollar figures from examination of budget statements. Appropriations for special mandates at the state level are one such example. Materials costs within a program may be another. Other testing resource costs can be converted into dollars, if necessary, by determining their shares of use in testing versus other activities and then prorating costs accordingly. The cost of school district personnel time is an example where this may be required. If a teacher spends ten percent of his or her time in assessment activities, an

equivalent share of salary and benefits could be attributed to testing in a cost analysis.

Many testing costs are not readily measured in dollars. Pupil time is one such cost, particularly at the elementary level where students have little or no "market" value in alternative settings such as the workplace; yet we do anticipate real costs to be associated with diverting pupils from other learning activities. The various policy costs listed in Figure 1 are also rather divorced from dollar equivalents. For example, we can only guess what the legislature might legislate, or the state education agency might develop and monitor, if they were not devoting time to minimum competency testing. Yet the fact that these offices devote resources to testing may be of interest in a comprehensive cost of testing investigation.

From a practical standpoint and for an initial inquiry, the evaluation of costs should begin with careful assessment of identified costs in their primary units. Teacher and pupil time should be observed in hours, along with time contributions of other professional and clerical staff. Budgeted figures for testing programs should be recorded in dollars, as should direct costs such as materials. The cataloguing of costs with appropriate values in this manner will provide basic data from which to analyze the costs of testing in a variety of conceivable ways.

Analysis of Costs--Some Hypothetical First Inquiries

We have referred generally to identifying and evaluating costs as they pertain to cost of testing inquiries and to their dependence of specific investigations which might be of interest. The term "costs of testing" conveys little meaning without some elaboration. An analysis of costs and testing must begin with a question or questions. The following examples

serve to illustrate the range of questions and the varying foci that might appear under the guise of "costs and testing." Each represents a distinct inquiry and the list is not exhaustive, but actual inquiry must begin with specific questions like these.

1. What is the total "cost" of all testing that is conducted in the schools? in a given state? nationwide?

2. What is the total "cost" of all testing in the classrooms of a school district that is mandated by state offices? by district offices?

3. What is the total "cost" of testing conducted for curricular purposes?

4. What are the costs associated with a particular type of test?

5. How do alternative means of designing and conducting statewide minimum competency tests for high school graduation compare on the basis of their costs? on the basis of their effectiveness and costs?

6. Should a state specify a competency test for use by all districts, or allow districts to develop their own tests within state guidelines?

7. What are the costs of compliance associated with a particular testing mandate? For a school? For a district? For a state as a whole?

8. How much testing should be incorporated in a 9th grade algebra curriculum? a 5th grade reading curriculum?

9. Should reading teachers in a particular context purchase end-of-level tests or develop their own tests?

While these questions are in some cases not pure inquiries into the costs of testing, each has significant cost components which could be assessed. Each involves specific units of analysis and implies the development of a unique inventory of costs. Beyond this, the nature of the questions asked will guide the evaluation and analysis of costs. The types of analysis which might be undertaken in regard to these questions range

from simplistic inventory processes to sophisticated cost effectiveness analysis and econometric analysis. We will describe several hypothetical types of analysis in reference to these questions, and in light of the previous discussion of testing cost analysis.

Analysis Using the Cost Inventory

The cost identification process yields information that can be useful for limited cost analysis and comparisons. While the cost inventory applied to a given test or test program is likely to be performed with subsequent and higher levels of analysis in mind, an inventory alone may be of interest. First, the inventory presents a map of the various costs associated with testing. Even this rudimentary level of knowledge about costs and testing is more than is often applied to issues of testing policies. Second, rough comparisons of testing programs can be made on the basis of cost inventories.The mere presence or absence of certain types of costs may be important considerations in testing decisions. For example, two testing strategies may appear to differ in cost only in their demands upon clerical time. If the relevant clerical staff is already fully engaged, this element of cost information could inform a decision about testing. And finally, the inventory itself provides a guide to subsequent questions in a cost analysis. Prior to the development of a cost inquiry, the investigator is often not fully aware of what questions will be important to his analysis.

Analysis Using "Total" Costs

While the cost inventory allows us to examine the types of costs involved with testing programs, the inventorying exercise does not lead to very precise assessments or comparisons unless the costs identified are also evaluated. As described above, a first approximation of total costs

can be obtained by estimating or recording an appropriate measure for each of the costs related to a test or testing program. This might yield a cost summary which looks like this hypothetical and very rudimentary example.

Figure 4

Sample Testing Cost Inventory and Evaluations:
Alternative School District Achievement Tests

Test A: Type of Cost	Estimated Level of Cost		
	Test A	Test B	Test C
Teacher time	100 hrs	100 hrs	50 hrs
Pupil time	2000 hrs	2000 hrs	2000 hrs
Clerical time	50 hrs	0 hrs	10
Materials	$1000	$1000	$500
Machine processing & fee	$2000	$2000	$3000

This example refers to a set of alternative hypothetical school district achievement tests. The information that comprises even this simple inventory illustrates the identification of the range of types of costs involved and the types of estimates (or calculations) that might be generated for each of those costs for each of the three tests. Before sketching a crude analysis of these figures, even this simple example raises questions as to how the cost identification and evaluation processes might be carried out in practice. These activities have been described more generally above, but specific comments can be directed to this example. Test data in this simple form does not normally exist in any one place in the records of schools or school districts. Given a particular test, such as a district achievement test, the investigator will necessarily have to survey individuals involved in the process to develop the needed information. Teachers are a likely source for much of the

information--they are certainly best qualified to provide estimates of pupil time and teacher time allocated to testing. Estimates of clerical time and materials and processing fees may be obtained from district business officers or perhaps from district testing coordinators. An important point is that a study of the costs of testing at this level involves directly accessing information from individuals at the school and district level.

The types of analysis that can be done with information shown in Figure 4 are quite limited but not insignificant. Test A and Test B offer a sort of comparison which was previously described. The two tests appear to involve very similar costs, but differ in that Test A requires 50 hours of clerical time and Test B requires none. If the tests were regarded equally as serving the information and other needs of the district, this analysis suggests that Test B would be preferred on the basis of costs. But if Test B was considered to be inferior to Test A (i.e., they differ in effectiveness), the decision is more complex. Nevertheless, at least the cost implications of a decision are illuminated in this comparison.

The comparison of Test A and Test C illustrates certain limitations of this type of analysis. The two tests vary considerably on each cost dimension, and the comparisons are highly inconclusive. While Test A is more costly than Test C in three of the identified areas, it is less costly in the remaining two. The utility of this comparison is constrained by the fact that the costs as presented are incommensurable. How does an hour of pupil time compare to a dollar of materials costs? and similar questions confound the analysis. While the example would allow us to say that Test B has lower costs than Test A, no such statement can be made for a comparison of Tests A and C. This type of analysis simply does not yield a single total cost figure from which such comparisons can be made.

Analysis Using Location of Costs

Tests vary in the degree to which their costs are distributed among various individuals and offices within the school and school policy-making systems. This was described in the general discussion of costs above by reference to the fact that programs frequently impose costs on entities far removed from the decision makers. The simple location of costs associated with a testing program or with a set of alternative programs may be a useful exercise. This would have limited value for an examination of, say, weekly curricular quizzes, since they are likely to involve only pupil and teacher time as significant costs in any configuration. But testing policies such as state mandates usually involve multiple levels in school policy-making and administrative systems--from the legislature and state education agency down to the pupil. In these programs, the costs are inevitably distributed across a variety of points within the total system. And alternative schemes may involve greatly differing distributions of costs regardless of their relative levels of costs.

Minimum competency tests for high school graduation offer a clear example of where cost location is important. A mandate might require districts to develop their own tests according to a set of guidelines, or it might simply specify a particular test or choice of tests. In the first case, a cost assessment would no doubt reveal a substantial level of costs imposed upon school districts (and a substantial total of costs due to the duplication of similar efforts across all districts). The second case might reveal high costs accruing to state offices for test development and implementation. Even without good measures of each cost, this locational type of cost information might benefit testing policy discussions.

Analysis Under the Economics of Information Paradigm

This model provides a theoretical model in which to consider the relationships between costs and outcomes of testing. As described, one of the suggestions of the paradigm is that ultimately both the benefits and costs of testing might be linked to the effectiveness of schooling. The resources applied to testing, whether in dollars spent or hours devoted to the processes by individuals, are resources that have alternative uses in the delivery of educational services. At the same time, testing provides benefits that might lead to enhanced delivery of those services and hence to greater pupil outcomes. So both the inputs (costs) of testing and the outcomes (benefits) could be reduced, at least theoretically, to their impacts on educational outcomes.

The general application of this paradigm to testing and schooling, however, presents numerous practical hurdles. In place of converting benefits and costs to dollar equivalents (which is required for cost-benefit analysis) this model would require each of the benefits and costs to be directly associated with its impact on pupil outcomes (e.g., pupil achievement, among others). This has direct analogies to the general inquiry into the effects of schooling over the past two decades which was in part fueled by the well known "Coleman Report" (1966). The subsequent studies of what factors contribute to schooling outcomes have probably done more to establish the difficulties of input-output analysis in education than to overcome them (see Cohn, 1979). Relating elements of testing to schooling outcomes will suffer from analogous and more severe shortcomings, because both the costs and benefits may be less concretely definable, and their links to pupil outcomes even more remote than the variables commonly employed in education production studies.

The economics of information paradigm does suggest to us at least one intriguing line of inquiry, however, despite some of its utilitarian shortcomings. Recent research into pupil learning in reading and mathematics has stressed the importance of pupil time in the learning process (Carroll, 1963; Wiley & Harnischfeger, 1974; Bloom, 1976), and the BTES Study reported by Denham and Lieberman (1980). These studies have a common quality in that they attempt to relate the amount of time devoted by pupils to learning activities, and/or the amount of time that pupils are actually engaged in such activities, to the performance of pupils on tests related to those learning activities. The most recent of these studies (BTES) builds elaborately on its predecessors. It offers not only comprehensive profiles of time use in second and fifth grade classrooms, but also estimates of the effects of time utilization on the outcomes measured. In the context of these types of estimates, pupil time devoted to testing may take on added meaning. Hours devoted to testing could be expressed in terms of their opportunity cost, i.e., hours not devoted to learning experiences. And these costs could be translated into estimated effects on learning using the BTES findings. Unfortunately, these data do not apply to high schools, and we do not yet have comparable studies at the secondary level. This approach would also have to acknowledge any incidental pupil learning that takes place because of testing activities, since testing is not, at this point, to be considered to be exclusively "down-time" from the pupil-learner's point of view.

This type of analysis might not add a substantial amount of information to a study of the total costs of testing, but if our inquiries suggest that pupil time is a substantial component of testing costs either generally or in specific contexts, then more thorough investigation of the importance of pupil time may be justified.

CONCLUSIONS

Costs and testing represent a recent merger of time-worn topics for education analysts. We know little about the costs of testing in the schools because few people have raised such questions. And we do not know much about the importance of such inquiries except that the proliferation of testing in recent years suggests that a look at the costs of testing may be overdue. In this chapter we present a global framework within which to think about the costs of testing--the economics of information paradigm. This framework suggests that we might think ultimately about the costs (and benefits) of testing in terms of their impact on instructional effectiveness in the schools--a construction at this time more appealing to theorists than to practical investigators. Even with this limitation, the paradigm subsumes the full range of questions and analyses that we might pose under the guise of costs and testing, and so serves a useful purpose.

Practical guidance to those interested in testing costs is offered in our discussion of the issues surrounding the identification and evaluation of testing costs. The results of these tasks form the basis for any type of cost analysis, and constitute necessary first steps for anyone presently pursuing questions of testing costs in the schools. Given both our collective inattention to the whole realm of the costs of testing, and given also the prerequisite nature of cost identification and evaluation, sharpening these notions and tending to their practical ramifications are top priorities as we contemplate future empirical studies of testing costs.

REFERENCES

Airasian, P.W., Madous, G., Pedulla, J., & Newton, K. Policy issues in minimal competency testing and a comparison of implementation models. Massachusetts Advisory Committee on High School Graduation Requirements, 1978.

Anderson, B.D. The costs of legislated minimal competency requirements. Background paper prepared for Minimal Competency Workshops sponsored by The Education Commission of the States and the National Institute of Education, 1977.

Anthony, R.N. Management accounting. Homewood, Ill: Irwin, 1964.

Bloom, B.S. Time and learning. American Psychologist, 1974, 29, 682-688.

Carroll, J.B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Cohn, E. The economics of education. Cambridge, MA: Ballinger, 1979.

Colin, E. The economics of education. Cambridge, MA: Ballinger, 1978.

Denham, C., & Liebermann, D.(Eds.) Time to learn. Washington, D.C. National Institute of Education, 1980.

Haller, E.J. Cost analysis for educational program evaluation. In W.J. Popham (Ed.) Evaluation in Education. Washington, D.C.: American Educational Research Association, 1974, Chap. 7

Hirshleifer, J., & Riley, J. The analytics of uncertainty and information: An expository survey. Journal of Economic Literature, 1979, 17, 1375-1421.

Horngren, C.T. Cost accounting. Englewood Cliffs, NJ: Prentice-Hall, 1967.

Levin, H.M. Cost effectiveness analysis in evaluation research, in M. Guttentag and E. Streuning (Eds.) Handbook of evaluation research, Volume 2, Beverly Hills, CA: Sage Publications, 1975, Chapter 5.

Levin, H.M. Educational performance standards: Image or substance, Journal of Educational Measurement, Winter 1978, 15, 4, 309-319.

Levin, H.M. Teacher certification and the economics of information. Educational Evaluation and Policy Analysis, July-August 1980, 2(4), 5-18.

Levin, H.M., & Seidman, W. Towards a meta cost effectiveness analysis of educational interventions. Report to the Educational Testing Service, Princeton, NJ, October 1981.

Lyon, C. New perspectives on school district research and evaluation. Paper prepared for the National Conference on Urban Education, St. Louis, MO, July 10-14, 1978.

Mansfield, E. Microeconomics, theory and application. New York, Norton, 1975.

McKean, R.N. Avoidance and enforcement costs in government regulation. St Louis Center for the Study of American Business, Wadington University, Working Paper No. 16, 1976.

Mishan, E.J. Cost benefit analysis. New York: Praeger, 1976.

Musgrave, R.A. Cost benefit analysis and the theory of public finance. Journal of Economic Literature, 1969, 7, 797-806.

Pipho, C. Update VIII: Minimum competency testing. Report No. 124, Education Commission of the States, 1979.

Resnick, D.P. Testing in America: A supportive environment. Phi Delta Kappan, May 1981, 631-634.

Schriber, P.E. Cost benefit analysis of comprehensive achievement monitoring for classroom evaluation. University of Massachusetts, Working Paper No. WP-18, February 1981.

Spence, M. Job market signaling. Quarterly Journal of Economics, 1973, 87, 355-374.

Sproull, L., & Zubrow, D. Standardized testing from the administrative perspective. Phi Delta Kappan, May 1981, 628-631.

Stigler, G.F. The economics of information. Journal of Political Economy, June 1961, 69, 3, 213-225.

Thomas, T.C., & Larson, M. Educational indicators and social policy. Cambridge: Aspen Institute for Humanistic Studies, 1976.

Wildavsky, A. Information as an organizational problem. Los Angeles: Graduate School of Management, University of California, Working Paper 1081, November 1981.

Wiley, D.E., & Harnishfeger, A. Explosion of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. Educational Researcher, 1974, 4, 7-12.